Institut d'études politiques de Paris

ECOLE DOCTORALE DE SCIENCES PO

Programme doctoral d'économie

Département d'économie

Doctorat en sciences économiques

THREE ESSAYS IN INFORMATION AND MECHANISM DESIGN

VICTOR AUGIAS

Thesis supervised by Jeanne Hagenbach, Directrice de recherche CNRS, Sciences Po, Département d'économie (UMR 8259)

Date of defense: May 24, 2023

Jury:

SARAH AUSTER (referee – rapporteure)

Associate Professor, Universität Bonn, Institute for microeconomics

JEANNE HAGENBACH (supervisor – directrice de thèse)

Directrice de recherche CNRS, Sciences Po, Département d'économie (UMR 8259)

FRÉDÉRIC KOESSLER (examiner – examinateur)

Directeur de recherche CNRS, GREGHEC (UMR 2959)

LAURENT MATHEVET (referee – rapporteur)

Professor of Economics, European University Institute, Department of economics

FRANZ OSTRIZEK (examiner – examinateur)

Assistant Professor, Sciences Po, Département d'économie

EDUARDO PEREZ-RICHET (examiner – examinateur)

Associate Professor, Sciences Po, Département d'économie

Note to the reader

The main language of this thesis is English. However, due to legal requirements, the general introduction is written in both English and French. The purpose of the introduction is to present the overarching issue connecting the three chapters and to outline their contributions to the existing literature. The acknowledgements are also written in both languages.

To facilitate the consultation of references, a specific bibliography is provided at the end of each chapter. The appendices are located at the end of the work and are themselves subdivided into chapters. In contrast to the main text, the bibliographic references cited in the appendices are all grouped together at the end of the appendices.

Note au lecteur

La langue principale utilisée pour la rédaction de cette thèse est l'anglais. Néanmoins, en raison de contraintes légales, l'introduction générale de cette thèse est rédigée en anglais et en français. L'objectif de l'introduction est de présenter la problématique d'ensemble reliant les trois chapitres et de mettre en évidence leurs contributions à la littérature existante. Les remerciements sont également écrits dans les deux langues.

Afin de faciliter la consultation des références, une bibliographie spécifique est ajoutée à la fin de chaque chapitre. Les annexes se trouvent à la fin de l'ouvrage et sont elles-mêmes subdivisées en chapitres. Contrairement au texte principal, les références bibliographiques citées dans les annexes sont toutes regroupées à la fin des annexes.

Acknowledgements

I want to express my deep gratitude to the people who contributed to the completion of this thesis.

First and foremost, I would like to thank my thesis advisor, Jeanne Hagenbach, for her unwavering support and trust during these past seven years. Initially as my Master's thesis supervisor at *ENS de Cachan*, and then throughout these five years of doctoral research at Sciences Po. You provided me with the ideal working conditions, that any doctoral student would dream of. You have skillfully guided my research with rigor and expertise, even in fields far removed from your own, while demonstrating exceptional kindness and patience. Everything I learned under your direction, both scientifically and personally, is invaluable.

This work owes a great deal to Eduardo Perez-Richet, who diligently followed the evolution of my research and played the role of a true "shadow director" throughout this thesis. I cannot thank you enough for your availability, patience, attentive listening, and always constructive and relevant feedback. Without your unfailing knowledge of literature, intellectual rigor, and impressive mastery of information economics models, I would never have been able to complete this dissertation. I am deeply grateful to you.

I want to warmly thank Sarah Auster, Frédéric Koessler, Laurent Mathevet, and Franz Ostrizek, who kindly agreed to be members of my jury, and especially Sarah and Laurent for agreeing to be the referees. I also want to thank Laurent for offering me the opportunity to spend the spring semester of 2022 at the Department of Economics at the European University Institute, in Florence. I am sincerely grateful for your insightful comments and suggestions during my stay there. Finally, I want to thank Franz for helping me overcome the job market with his insightful advice and compassionate listening. I also want to express my appreciation to Sarah for welcoming me during my (brief) visit to the University of Bonn, which will be my new academic home for the years to come.

I would also like to express my gratitude towards my co-authors, Alexis and Daniel. This thesis is yours as well. As theorists become increasingly rare among academic economists, I can only appreciate the luck I had to share my passion with the only other microeconomist in my cohort of PhD students at Sciences Po, Daniel, in whom I found not only a brilliant work companion but also a true friend. I will sadly miss our Friday night dinners at Bangkok Street Food, drinks in Belleville, and pints at Supercoin. Alexis, although our meeting came later, it was no less

intense. I look forward to visiting you in your Oxford lair to explore the uncharted territories of mechanism design and of the London underground jazz scene.

A huge thank you to my fellow PhD students at Sciences Po. This thesis owes a lot to all of you. First, my illustrious predecessors, Aseem, Alaïs, Charles, Elisa, Edoardo, Oliver, Nicolò, Ludovic, Sophie, and Vladimir, who welcomed me when I arrived in the *combles*, in 2018. Then, those with whom I shared the majority of my thesis years, Ariane, Clemens, Diego, Edgard, Gustave, Jérôme, Juan, Léonard, Marco, Mylène, Pauline, Pierre Coster, Pierre Villeudieu, Samuel, Ségal, Sophia, Stefan, Victor, and Zydney. And finally, my former students who became PhD students, Aurélien, Claudius, Constance, Naomi, Nathaniel, Olivia, Rémi, and Riddhi and Valentin. I would also like to thank all the PhD students I had the opportunity to meet at the EUI during my visit for their warm welcome. Foremost among them are Grégory and Thomas, who greatly facilitated my integration into the department.

Although it takes up a significant part, a PhD student's life is not limited to research. I am immensely grateful to the people who have accompanied me throughout this thesis. First and foremost, my grandparents, my parents and my sister for their love and unfailing support throughout my studies. My long-time southern friends, Adrien, Hugo, Maxime, Timothée and Yohan, my friends from the *Magistère d'économie de Paris 1*, Amandine, Antoine, Florian, Gabrielle, Guillaume, Julie, Lotfi, Marie, Marie-Sophie, Maxime, Philippe, Pierre, Sarah and Simon, as well as the team from the 18th *arrondissement*, now scattered all over Paris, Azalaïs, Flore, Luce and Majdi. And finally, Dominique and Francis, for their support and all those comforting Sunday lunches.

Finally, I would like to thank Anna for these six years of common life. You have supported me in all the moments, the highs and the lows (even the very low ones). I will never thank you enough for all that you accepted to endure so that I could finish this dissertation. You followed every step of it. You took a sincere interest in my work when I no longer saw any meaning in it. You knew how to motivate me in moments when I was close to giving up. You removed my doubts during the most stressful moments. You have unconditionally accepted that I cross the Rhine alone to pursue my career. Without you, I would never have been able to put the final touches on this manuscript. This thesis is dedicated to you.

Remerciements

Je tiens à exprimer ma profonde gratitude envers toutes les personnes qui ont contribué à la réalisation de cette thèse.

Je tiens tout d'abord à remercier ma directrice de thèse, Jeanne Hagenbach, pour son soutien inconditionnel et sa confiance totale lors de ces sept dernières années. Tout d'abord en tant que directrice de mémoire à l'ENS de Cachan, puis, au cours de ces cinq années de thèse à Sciences Po. Tu m'as offert les conditions de travail dont tout doctorant rêverait. Tu as su guider mes recherches avec rigueur et expertise, même dans des domaines éloignés des tiens, tout en faisant preuve d'une bienveillance et d'une patience à toute épreuve. Tout ce que j'ai appris sous ta direction, tant sur le plan scientifique qu'humain, est inestimable.

Ce travail doit énormément à Eduardo Perez-Richet, qui a assidûment suivi l'évolution de mes travaux et a joué le rôle d'un véritable directeur de l'ombre tout au long de cette thèse. Je ne saurais assez te remercier pour ta disponibilité, ta patience, ton écoute et tes commentaires toujours aussi constructifs que pertinents. Je n'aurais jamais pu venir à bout de cette thèse si tu ne m'avais pas partagé ta connaissance sans faille de la littérature, ta rigueur intellectuelle, ainsi que ta maîtrise impressionnante des modèles d'économie de l'information. Je t'en suis profondément reconnaissant.

Je tiens à remercier chaleureusement Sarah Auster, Frédéric Koessler, Laurent Mathevet et Franz Ostrizek, qui ont gentiment accepté d'être membres de mon jury, et plus particulièrement Sarah et Laurent de bien avoir voulu en être les rapporteurs. Je tiens également à remercier Laurent pour m'avoir offert l'opportunité de passer le semestre de printemps 2022 au sein du département d'économie de l'Institut Universitaire Européen, à Florence. Je te remercie sincèrement pour tes commentaires et tes suggestions judicieux lors de ce séjour. Je tiens également à remercier Franz, pour m'avoir aidée à surmonter l'épreuve du *job market* grâce à ses conseils avisés et son écoute bienveillante. Finalement, je souhaite également remercier Sarah pour son accueil lors de ma (courte) visite à l'université de Bonn, qui sera mon nouveau département pour les années à venir.

Je tiens également à exprimer ma gratitude envers mes co-auteurs, Alexis et Daniel. Cette thèse est aussi la vôtre. Alors que les théoriciens se font de plus en plus rares parmi les économistes académiques, je ne peux qu'apprécier la chance que j'ai eue de partager ma passion avec le seul autre microéconomiste de ma cohorte de doctorants à Sciences Po, Daniel, en qui j'ai trouvé non seulement un

brillant compagnon de travail mais aussi un véritable ami. Nos dîners du vendredi soir chez Bangkok Street Food, nos apéros à Belleville et nos pintes au Supercoin me manqueront cruellement. Alexis, bien que notre rencontre ait eu lieu plus tard, elle n'en a pas été moins intense. J'ai hâte de te rendre visite dans ton repaire d'Oxford pour découvrir des territoires inexplorés du *mechanism design* et de la scène jazz *underground* londonienne.

Un grand merci à mes compagnons de fortune des "combles". Cette thèse vous doit beaucoup. D'abord, mes illustres prédécesseurs, Aseem, Alaïs, Charles, Elisa, Edoardo, Oliver, Nicolò, Ludovic, Sophie, et Vladimir, qui m'ont accueilli à mon arrivée en 2018. Ensuite, ceux avec qui j'ai partagé la majorité de mes années de thèse, Ariane, Clemens, Diego, Edgard, Gustave, Jérôme, Juan, Léonard, Marco, Mylène, Pauline, Pierre Coster, Pierre Villeudieu, Samuel, Ségal, Sophia, Stefan, Victor, et Zydney. Et enfin, mes anciens étudiants devenus doctorants, Aurélien, Claudius, Constance, Naomi, Nathaniel, Olivia, Rémi, et Riddhi et Valentin. Je tiens également à remercier tous les doctorants que j'ai eu l'occasion de rencontrer à l'EUI lors de ma visite pour leur accueil chaleureux. Au premier rang desquels Grégory et Thomas, qui ont grandement facilité mon intégration au sein du département.

Même si elle prend une place conséquente, la vie d'un thésard ne se limite pas à la recherche. Je suis immensément reconnaissant aux personnes qui m'ont accompagné tout au long de cette thèse. Tout d'abord, mes grand-parents, mes parents et ma sœur pour leur amour et leur soutien indéfectible tout au long de mes études. Mes amis sudistes de longue date, Adrien, Hugo, Maxime, Timothée et Yohan, mes amis du Magistère d'économie de Paris 1, Amandine, Antoine, Florian, Gabrielle, Guillaume, Julie, Lotfi, Marie, Marie-Sophie, Maxime, Philippe, Pierre, Sarah et Simon, ainsi que l'équipe du 18ème arrondissement, maintenant dispersée aux quatre coins de Paris, Azalaïs, Flore, Luce et Majdi. Et enfin, Dominique et Francis, pour leur soutien et tous ces réconfortants déjeuners dominicaux.

Enfin, je voudrais remercier Anna pour ces six années de vie commune. Tu m'as soutenu dans tous les moments, les hauts comme les bas (voire les très bas). Je ne te remercierai jamais assez pour tout ce que tu as accepté d'endurer afin que je puisse terminer cette thèse. Tu en as suivi toutes les étapes. Tu t'es sincèrement intéressé à mon travail lorsque je n'y voyais plus de sens. Tu as su me motiver dans les moments où j'étais sur le point d'abandonner. Tu as levé mes doutes dans les moments les plus stressants. Tu as accepté sans condition que je traverse seul le Rhin pour poursuivre ma carrière. Sans toi, je n'aurais jamais pu mettre le point final à ce manuscrit. Cette thèse t'est dédiée.

Contents

List of Figures

Introduction

Information structures, from which economic agents learn and form their beliefs, are largely *endogenous*. They are often *designed* by individuals or organizations to serve specific objectives. When information is disclosed strategically, it becomes a powerful instrument for providing incentives. By shaping agents' beliefs, it influences their actions and, as a result, the outcomes of economic, social, and political interactions. The recent increase in data production, combined with advancements in data processing, has transformed information into a pervasive instrument of *persuasion*. In modern economies, these structures manifest themselves, for example, in the form of algorithms, such as recommendation systems used by streaming (Che and Hörner, 2018) and matching platforms (Romanyuk and Smolin, 2019), or predictive algorithms employed by police forces and law enforcement agencies (Hernández and Neeman, 2022; Ichihashi, 2023). Other notable examples include rating (Saeedi and Shourideh, 2020; Vellodi, 2018), grading (Boleslavsky and Cotton, 2015), certification (Zapechelnyuk, 2020), scoring (Ball, 2019), and performance measurement systems (Georgiadis and Szentes, 2020; Ostrizek, 2022), or even financial stress tests (Goldstein and Leitner, 2018; Inostroza, 2019; Orlov, Zryumov, and Skrzypacz, 2021). This dissertation studies how information structures are and should be designed to achieve desirable social goals, as well as the constraints limiting the use of information as an incentive tool.

The idea that information is endogenous is firmly rooted in the economics and game theory literature. Information can be strategically revealed or concealed through a variety of communication protocols, which depend on technological and institutional contexts. In the field of strategic communication, the contribution of Kamenica and Gentzkow (2011) marks a significant turning point. Traditionally, economists classify communication protocols into three categories depending on whether the transmission of information is costly (Spence, 1973), non-costly (Crawford and Sobel, 1982), or certifiable (Milgrom, 1981; Grossman, 1981). The innovation brought by Kamenica and Gentzkow, in comparison to these three communication modes, is to endow the information designer with *full commitment power*.[1] This assumption means that the information designer determines the

[1]Information design obviously dues its name to mechanism design, where a similar commitment assumption is made. Mechanism design concentrates on identifying the feasible set of institutions (i.e., the rules of the game) and how to optimally design them according to some social criterion, assuming the designer possesses full commitment power over the game that will be played by agents, and the players' information structure is fixed. For instance, auction design consists in determining

receiver's information structure *ex-ante*, that is, at the moment when the realization of the state of the world is not yet known, and *cannot deviate from it* when the state of the world is realized. Much of the literature has sought to provide microeconomic foundations for the the information designer's commitment power assumption, which can be justified on technological, reputational, and credibility grounds (see, for example, Best and Quigley, 2017; Mathevet, Pearce, and Stacchetti, 2022; Lipnowski, Ravid, and Shishkin, 2022; Lin and Liu, 2022, as well as the references cited by the authors). For instance, once a technological company has designed the recommendation algorithm used to provide information to its platform users, it is difficult for the company to deviate from it in the short term without incurring prohibitive costs. Similarly, an institution responsible for maintaining the stability of the financial system has every interest in establishing transparent and unalterable rules to maintain its reputation and, consequently, the credibility of its *stress tests*.

From a theoretical standpoint, particular attention has been paid to the questions of feasibility and optimality: to what extent does information design allow for influencing behaviors, and given the range of possibilities, what is the optimal information structure from the designer's perspective? Kamenica and Gentzkow address these two questions by proposing a characterization of the payoffs achievable through information design and a method for systematically obtaining the optimal structure for the designer, in the case where *only one agent* receives the information. These results have subsequently been extended in various directions, for instance, when information design is costly (Gentzkow and Kamenica, 2014), when multiple information designers are in competition (Gentzkow and Kamenica, 2017), or to dynamic environments (Ely, 2017), as well as to the broader framework of incomplete information games (Bergemann and Morris, 2016; Mathevet, Perego, and Taneva, 2020).[2] This body of work demonstrates the vast potential offered by information design in terms of implementable receivers' behaviors and, consequently, welfare outcomes resulting from receivers' strategic interactions. Indeed, the only constraint imposed by information is of a statistical nature: the designer can only induce beliefs that are consistent with Bayes' rule and, therefore, the receivers' prior beliefs.[3] Consequently, all behaviors that are compatible with

the set of implementable allocations and how to maximize the revenue or social welfare from the auction, the bidders' information structure being given. In contrast, information design maintains players' incentives constant, focusing solely on manipulating the game's information structure.

[2]For a comprehensive presentation of these extensions, we refer to the literature reviews of Bergemann and Morris (2019) and Kamenica (2019).

[3]This condition, called *Bayes plausibility* by Kamenica and Gentzkow, is sufficient when the designer aims to influence the behavior of a single agent. When multiple agents interact strategically after information disclosure, the set of beliefs that can be induced by the designer must also be compatible with the higher-order beliefs of the players (Mathevet et al., 2020).

Bayesian updating are theoretically attainable for an information designer with full commitment power.[4] For instance, in a more concrete setting, Bergemann, Brooks, and Morris (2015) show that, for monopolistic markets, an adequate design of the seller's information structure regarding consumers' marginal willingness to pay could lead to situations where the gains from trade are entirely captured by the seller, as well as situations where they are entirely captured by the consumers, and all intermediate scenarios, provided that the revealed information does not alter the seller's profit compared to the situation where no information would be revealed. Thus, it is theoretically possible for a benevolent information designer, such as a regulatory authority, to distribute the entire surplus to consumers by revealing only information to the monopoly.

Although recent research has led to significant theoretical and applied advancements, outlining the possibilities offered by information alone and describing the properties of optimal information structures in various contexts, the design of information structures remains subject to numerous practical constraints, which are only partially accounted for by the theory. As Kamenica, Kim, and Zapechelnyuk (2021) emphasize in their recent editorial:

> "The basic theory [of information design] makes a number of assumptions, which are sufficiently plausible in many contexts and have enabled various novel insights. Yet, we think that a more flexible approach that relaxes these assumptions would significantly enhance the applicability of the theory. Here we focus on two of the assumptions. First, the receivers are the standard rational players who maximize their expected utility and *make Bayesian inferences*. Second, there are *few or no constraints on feasible information structures (signals, experiments)*."

The first two chapters of this dissertation aim precisely to integrate into the theory an important type of constraint, yet unexplored by the literature, as well as to relax the assumption of Bayesian belief updating by the receiver.

Information design and upstream investment incentives. Agents can change their behavior in response to how information is structured *before it is even revealed*. These *upstream* strategic behaviors limit the set of information structures that the designer can use, as he must ensure that they are *compatible with the*

[4]Importantly, this conclusion rests entirely on the commitment assumption. When the sender has no commitment power (Crawford and Sobel, 1982), his disclosure strategy must be consistent with his incentive constraints additionally to the Bayesian update condition. This implies that the meaning of the signals generated by the sender's strategy is determined *in equilibrium*, unlike the case with commitment where the meaning of the signals is *objectively determined*.

players' incentives upstream of information production. For example, a large number of goods and services are subject to *tests*, set up by public regulators, to determine whether they can be approved or receive certification (such as quality labels or green labels). However, the way tests are designed has an upstream impact on firms' incentives to participate in those tests (Rosar, 2017; Harbaugh and Rasmusen, 2018), or on fraud and falsification behaviors (Perez-Richet and Skreta, 2022a,b). Therefore, it is crucial for regulators to take these incentives into account when designing the tests. In chapter 1, co-authored with Eduardo Perez-Richet, we examine another type of upstream strategic behavior: *productive investments*.[5] We develop an information design model where the state of the world can be *transformed* upstream by an agent at a certain cost. The designer's goal is to act only if the *final* state of the world is sufficiently high, while the agent's goal is for the designer to act. This model is particularly applicable to *selection* situations. For example, in the case of tests, the state of the world corresponds to the quality of the product a firm is seeking to certify, and the designer is a regulator deciding whether or not to approve the product. The information structure then corresponds to a test revealing information to the regulator about the product's quality after the firm's investment. The regulator's goal is to acquire the best possible information for making their decision, while the firm seeks to maximize the probability of their product being approved by the regulator. Our main result shows that the optimal information structures for the designer are *binary* and *deterministic*. In the case of the regulator, this corresponds to setting a *threshold* of quality beyond which a signal indicating to approve the product is generated by the test and below which a signal indicating to reject it is generated. This result establishes a theoretical foundation for "pass or fail" selection rules, widely used in practice, and surprisingly shows that it is suboptimal to introduce randomness into the information structure when strategic investments can be made upstream of information production, contrary to the cases of participation and falsification.

Information design and motivated belief updating. Setting aside strategic behaviors upstream, the receiver's downstream behavior when receiving information can be influenced by how they *choose to interpret* the information. Although agents learn from the information they receive, a vast literature in behavioral and experimental economics shows that they do not systematically do so as a statistician would, as prescribed by Bayes' rule, but rather let their preferences

[5] Although this chapter is formulated in terms of a resource allocation problem, we establish an equivalence result between this *allocation mechanism design* problem and the *information structure design* problem in the section 1.5.3.

dictate, to some extent, how they form their beliefs (see notably, Bénabou and Tirole, 2016, p. 150, Epley and Gilovich, 2016, or Benjamin, 2019, section 9). We show that this type of *motivated* reasoning (according to the term of Kunda, 1990) is likely to affect the effectiveness of information as an incentive instrument and how information should be revealed compared to the Bayesian case. We base our motivated beliefs model on the one proposed by Caplin and Leahy (2019): after observing an informative signal, the receiver forms beliefs by comparing their anticipatory value and their psychological cost. This modeling assumption reflects, on the one hand, that beliefs have an *intrinsic value* for the agent. For example, it has been shown that individuals associate utility with beliefs about their self-image or future prospects, such as thinking they are better than others, being in good health, moving upward in the social ladder, etc. (Bénabou and Tirole, 2016). On the other hand, distorting one's beliefs from Bayesian beliefs is *psychologically costly*. Evidence shows that agents employ sophisticated and costly mental strategies to protect or attain desirable beliefs, such as manipulating their own memory (see Bénabou, 2015; Bénabou and Tirole, 2016; Hagenbach and Koessler, 2022, and the references therein) or avoiding useful information (Golman, Hagmann, and Loewenstein, 2017), such as reliable tests for detecting serious illnesses (Oster, Shoulson, and Dorsey, 2013). In our model, the receiver's beliefs thus depend on their preferences as a result of motivated belief updating, and overweight the states of the world associated with the highest potential payoff of the receiver. We show that distortion of the receiver's beliefs leads to distortion of his behavior. Compared to a Bayesian individual, he prefers actions associated with the highest payoff and the highest variability of payoff, and if one action induces the highest payoff but another induces the highest variability of payoff, the preference for one or the other depends on the magnitude of the receiver's belief distortion cost. Consequently, *the efficacy of information as an incentive tool may vary according to individuals' material stakes*: persuasion is more effective, compared to the Bayesian case, when aimed at encouraging risky but potentially highly profitable behavior, and less effective when targeting more cautious behavior. We illustrate this result with applications, showing why information campaigns often prove ineffective in stimulating increased investment in preventive health treatments and how financial advisors can benefit from their clients' overoptimistic beliefs. We also show that strategic disclosure of information to voters with heterogeneous partisan preferences can lead to belief polarization within the electorate.

The distributional impacts of information design. Chapter 3 addresses a question of a different nature than the first two chapters. The literature has mainly focused on characterizing the payoffs achievable through information design from an *aggregate* perspective. On the contrary, the *distributional* effects of information design have been neglected, i.e., how the revelation of information affects different types of agents.[6] We examine this issue in the specific context of monopolistic markets, where the of the distribution of gains from trade generated by information is a first-order issue. Indeed, consumers are constantly leaving traces of their identity on the Internet, whether through their activity on social networks, their use of search engines, or their online purchases. The data generated and collected on consumers has become very valuable, as it allows companies to *segment* consumers, i.e., to pool them into subgroups based on observable characteristics and to price discriminate according to the subgroup. As mentioned earlier in the introduction, Bergemann et al. (2015) show that revealing information about consumers' marginal willingness to pay to the seller can induce a wide variety of outcomes in terms of segmentation and social surplus. In particular, it is always possible to segment the consumer population in a way that allocates the entirety of the gains from trade to them. To achieve this, segments can be created that pool consumers with high marginal willingness to pay together with those having a low marginal willingness to pay. Such pooling forces the monopolist to charge lower prices on those segments[7] and, as a result, to make consumers with high marginal willingness to pay benefit from lower prices. However, a significant drawback of this type of segmentation is that if the marginal willingness to pay and consumers' wealth are positively correlated, segmentations that maximize consumer surplus tend to *favor the wealthiest consumers*. How can the seller's information be designed to benefit the poorest consumers? We answer this question by studying a model of market segmentation in a monopolistic setting assuming a redistributive objective. We show that redistributive-optimal segmentations always generate Pareto-efficient allocations but may require granting a strictly positive share of the surplus to the seller. We identify the set of markets for which redistributive segmentation involves leaving a rent to the monopoly. We also develop a procedure for constructing the redistributive-optimal segmentation and show that, when the correlation between willingness to pay and consumer wealth is sufficiently high, the redistributive-optimal segmentation divides consumers

[6] A first foray in this direction is the contribution of Doval and Smolin (2021).

[7] Otherwise, the monopoly would decrease too much the extensive margin compared to the intensive margin, by excluding too many buyers from consumption, which would decrease its profits.

into contiguous segments: consumers are ranked and grouped in an increasing manner based on their willingness to pay. In this way, poorer consumers benefit from lower prices than the wealthier ones. Interestingly, this contrasts sharply with segmentations that maximize consumer surplus without distributional concerns, which group wealthy and poor consumers together in order to provide lower prices to the wealthier ones.

Bibliography

Ball, I. (2019): "Scoring Strategic Agents," *arXiv*.

Bénabou, R. (2015): "The Economics of Motivated Beliefs," *Revue d'économie politique*, 125, 665–685.

Bénabou, R. and J. Tirole (2016): "Mindful Economics: The Production, Consumption, and Value of Beliefs," *Journal of Economic Perspectives*, 30, 141–164.

Benjamin, D. J. (2019): "Errors in probabilistic reasoning and judgment biases," in *Handbook of Behavioral Economics: Applications and Foundations 2*, Elsevier B.V., vol. 2, chap. 2, 69–186.

Bergemann, D., B. Brooks, and S. Morris (2015): "The Limits of Price Discrimination," *American Economic Review*, 105, 921–57.

Bergemann, D. and S. Morris (2016): "Bayes correlated equilibrium and the comparison of information structures in games," *Theoretical Economics*, 11, 487–522.

——— (2019): "Information Design: A Unified Perspective," *Journal of Economic Literature*, 57, 44–95.

Best, J. and D. Quigley (2017): "Persuasion for the Long Run," *SSRN Electronic Journal*.

Boleslavsky, R. and C. Cotton (2015): "Grading Standards and Education Quality," *American Economic Journal: Microeconomics*, 7, 248–279.

Caplin, A. and J. Leahy (2019): "Wishful Thinking," *NBER Working Paper Series*.

Che, Y.-K. and J. Hörner (2018): "Recommender Systems as Mechanisms for Social Learning," *Quarterly Journal of Economics*, 133, 871–925.

Crawford, V. P. and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50, 1431–1451.

Doval, L. and A. Smolin (2021): "Information Payoffs: An Interim Perspective," *Working paper*.

Ely, J. C. (2017): "Beeps," *American Economic Review*, 107, 31–53.

Epley, N. and T. Gilovich (2016): "The Mechanics of Motivated Reasoning," *Journal of Economic Perspectives*, 30, 133–140.

Gentzkow, M. and E. Kamenica (2014): "Costly Persuasion," *American Economic Review: Papers & Proceedings*, 104, 457–462.

——— (2017): "Competition in Persuasion," *Review of Economic Studies*, 84, 300–322.

Georgiadis, G. and B. Szentes (2020): "Optimal Monitoring Design," *Econometrica*, 88, 2075–2107.

Goldstein, I. and Y. Leitner (2018): "Stress tests and information disclosure," *Journal of Economic Theory*, 177, 34–69.

Golman, R., D. Hagmann, and G. Loewenstein (2017): "Information Avoidance," *Journal of Economic Literature*, 55, 96–135.

Grossman, S. J. (1981): "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics*, 24, 461–483.

Hagenbach, J. and F. Koessler (2022): "Selective memory of a psychological agent," *European Economic Review*, 142, 104012.

Harbaugh, R. and E. Rasmusen (2018): "Coarse Grades: Informing the Public by Withholding Information," *American Economic Journal: Microeconomics*, 10, 210–235.

Hernández, P. and Z. Neeman (2022): "How Bayesian Persuasion Can Help Reduce Illegal Parking and Other Socially Undesirable Behavior," *American Economic Journal: Microeconomics*, 14, 186–215.

Ichihashi, S. (2023): "Privacy, Transparency, and Policing," *Working paper*.

Inostroza, N. (2019): "Persuading Multiple Audiences: An Information Design Approach to Banking Regulation," *SSRN Electronic Journal*.

Kamenica, E. (2019): "Bayesian Persuasion and Information Design," *Annual Review of Economics*, 11, 249–272.

Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

Kamenica, E., K. Kim, and A. Zapechelnyuk (2021): "Bayesian persuasion and information design: perspectives and open issues," *Economic Theory*, 72, 701–704.

Kunda, Z. (1990): "The case for motivated reasoning." *Psychological Bulletin*, 108, 480–498.

Lin, X. and C. Liu (2022): "Credible Persuasion," in *Proceedings of the 23rd ACM Conference on Economics and Computation*, New York, NY, USA: ACM, 469–469.

Lipnowski, E., D. Ravid, and D. Shishkin (2022): "Persuasion via Weak Institutions," *Journal of Political Economy*, 130, 2705–2730.

Mathevet, L., D. Pearce, and E. Stacchetti (2022): "Reputation for a degree of honesty," *Working paper*.

Mathevet, L., J. Perego, and I. Taneva (2020): "On Information Design in Games," *Journal of Political Economy*, 128, 1370–1404.

Milgrom, P. R. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12, 380.

Orlov, D., P. Zryumov, and A. Skrzypacz (2021): "The Design of Macro-Prudential Stress Tests," *SSRN Electronic Journal*.

Oster, E., I. Shoulson, and E. R. Dorsey (2013): "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease," *American Economic Review*, 103, 804–830.

Ostrizek, F. (2022): "Vague by Design: Performance Evaluation and Learning from Wages," *Working paper*.

Perez-Richet, E. and V. Skreta (2022a): "Fraud-proof non-market allocation mechanisms," *Working paper*.

——— (2022b): "Test Design Under Falsification," *Econometrica*, 90, 1109–1142.

Romanyuk, G. and A. Smolin (2019): "Cream Skimming and Information Design in Matching Markets," *American Economic Journal: Microeconomics*, 11, 250–276.

ROSAR, F. (2017): "Test design under voluntary participation," *Games and Economic Behavior*, 104, 632–655.

SAEEDI, M. AND A. SHOURIDEH (2020): "Optimal Rating Design," *arXiv*.

SPENCE, M. (1973): "Job Market Signaling," *Quarterly Journal of Economics*, 87, 355–374.

VELLODI, N. (2018): "Ratings Design and Barriers to Entry," *SSRN Electronic Journal*.

ZAPECHELNYUK, A. (2020): "Optimal Quality Certification," *American Economic Review: Insights*, 2, 161–176.

Introduction en français

Les structures informationnelles, à partir desquelles les agents économiques apprennent et forment leurs croyances, sont en grande partie *endogènes*. Elles sont souvent *conçues* par des individus ou des organisations afin de servir des objectifs spécifiques. Lorsque l'information est divulguée de manière stratégique, elle devient un puissant levier d'incitation. En influençant les croyances des agents, elle modifie leurs actions et, par conséquent, les résultats des interactions économiques, sociales et politiques. L'accroissement récent de la production de données, associé à l'amélioration de leur traitement, a transformé l'information en un instrument de *persuasion* omniprésent. Dans les économies modernes, les structures d'information se matérialisent, par exemple, sous la forme d'algorithmes, tels que les systèmes de recommandation utilisés par les plateformes de *streaming* (Che and Hörner, 2018) et d'appariement (Romanyuk and Smolin, 2019), ou les algorithmes prédictifs employés par les forces de police et les services de maintien de l'ordre (Hernández and Neeman, 2022; Ichihashi, 2023). Les systèmes de notation (Saeedi and Shourideh, 2020; Vellodi, 2018), d'évaluation (Boleslavsky and Cotton, 2015), de certification (Zapechelnyuk, 2020), d'attribution de scores (Ball, 2019), de mesure de la performance (Georgiadis and Szentes, 2020; Ostrizek, 2022), ou encore les *stress tests* auxquels sont soumises les institutions financières (Goldstein and Leitner, 2018; Inostroza, 2019; Orlov, Zryumov, and Skrzypacz, 2021) en sont d'autres exemples saillants. Cette thèse vise à étudier la manière dont les structures d'information sont et devraient être conçues afin d'atteindre des objectifs sociaux souhaitables, ainsi que les contraintes limitant l'utilisation de l'information en tant qu'outil d'incitation.

L'idée que l'information est endogène est bien établie dans la littérature en économie et en théorie des jeux. L'information peut être révélée ou dissimulée de manière stratégique via différents protocoles de communication, qui varient selon le contexte technologique et institutionnel. Dans le domaine de la communication stratégique, la contribution de Kamenica and Gentzkow (2011) marque un tournant significatif. Traditionnellement, les économistes classent la communication en trois catégories selon que l'information est coûteuse (Spence, 1973), non coûteuse (Crawford and Sobel, 1982) ou certifiable (Milgrom, 1981; Grossman, 1981). L'innovation apportée par Kamenica and Gentzkow, par rapport à ces trois modes de communication, consiste à doter le concepteur d'information d'un *pouvoir*

d'engagement total[8]. Cela signifie que le concepteur détermine la structure d'information du récepteur *ex-ante*, c'est à dire à l'instant où la réalisation de l'état du monde n'est pas encore connue, et *ne peut pas la modifier* lorsque l'état du monde devient connu. Une grande partie de la littérature s'est efforcée à donner des fondements microéconomiques à l'hypothèse du *pouvoir d'engagement* du concepteur d'information, qui peut être justifiée sur des bases technologiques, de réputation et de crédibilité (voir par exemple Best and Quigley, 2017; Mathevet, Pearce, and Stacchetti, 2022; Lipnowski, Ravid, and Shishkin, 2022; Lin and Liu, 2022, ainsi que les références citées par les auteurs). Par exemple, une fois qu'une entreprise technologique a conçu l'algorithme de recommandation utilisé pour fournir de l'information aux utilisateurs de sa plateforme, il lui est difficile de s'en écarter à court terme sans encourir un coût prohibitif. De même, une institution garante de la stabilité du système financier a tout intérêt à établir des règles transparentes et intangibles pour maintenir sa réputation et, par conséquent, la crédibilité de ses *stress tests*.

D'un point de vue théorique, une attention particulière a également été portée aux questions de la faisabilité et de l'optimalité : dans quelle mesure la conception d'information permet-elle d'influencer les comportements et, étant donné l'ensemble des possibles, quelle est la structure d'information optimale du point de vue du concepteur ? Kamenica and Gentzkow ont répondu à ces deux questions en caractérisant l'ensemble des gains atteingnables par la conception d'information, ainsi qu'une méthode permettant d'obtenir de manière systématique la structure optimale pour le concepteur, dans le cas où *un seul agent* reçoit l'information. Ces résultats ont par la suite été étendus, par exemple, aux cas où la conception de l'information est coûteuse (Gentzkow and Kamenica, 2014), où plusieurs concepteurs d'information sont en compétition (Gentzkow and Kamenica, 2017), aux environnements dynamiques (Ely, 2017), ainsi qu'au cadre plus général des jeux à information incomplète (Bergemann and Morris, 2016; Mathevet, Perego, and Taneva, 2020)[9]. L'ensemble de ces travaux montre l'ampleur des possibilités offertes

[8]Le terme "conception de l'information" tire évidemment son nom de la "conception de mécanismes", où une hypothèse d'engagement similaire est faite. La conception de mécanismes se concentre sur l'identification de l'ensemble réalisable des institutions (c'est-à-dire des règles du jeu) et comment les concevoir de manière optimale selon un critère social particulier, en supposant que le concepteur a un plein pouvoir d'engagement sur le jeu auquel les agents participeront et que la structure d'information des joueurs est fixe. Par exemple, la conception d'enchères consiste à déterminer l'ensemble des allocations réalisables et à maximiser les revenus ou le bien-être social issus de l'enchère, la structure d'information des enchérisseurs étant donnée. En revanche, la conception de l'information maintient les incitations des joueurs constantes, et se concentre uniquement sur la manipulation de la structure d'information du jeu.

[9]Pour une présentation exhaustive de ces extensions, nous renvoyons aux revues de la littérature de Bergemann and Morris (2019) et Kamenica (2019).

par la conception de structures d'information en termes de comportements des récepteurs et, par conséquent, de bien-être résultant de leurs interactions. En effet, l'unique contrainte liée à l'information est d'ordre statistique : les seules croyances que le concepteur peut induire chez un récepteur rationnel sont celles qui sont compatibles avec la règle de Bayes et, donc, la croyance *a priori* du récepteur[10]. Par conséquent, tous les comportements pouvant découler d'une mise à jour bayésienne des croyances sont théoriquement atteignables pour un concepteur d'information disposant d'un pouvoir d'engagement total[11]. Par exemple, dans un contexte plus concret, Bergemann, Brooks, and Morris (2015) ont démontré que, pour les marchés monopolistiques, une conception adéquate de la structure d'information du vendeur concernant la disposition marginale à payer des consommateurs peut aussi bien conduire aux situations où les gains à l'échange sont entièrement captés par le vendeur, qu'à celle où ils sont entièrement captés par les consommateurs, ainsi qu'à toutes les situations intermédiaires, à condition que l'information révélée n'altère pas le profit du vendeur par rapport à la situation où il n'obtiendrait aucune information. Ainsi, il est théoriquement possible pour un concepteur d'information bienveillant, tel qu'une autorité de régulation, de distribuer tout le surplus aux consommateurs en révélant seulement de l'information au monopole.

Bien que les recherches récentes aient conduit à d'importantes avancées théoriques et appliquées, délimitant les possibilités offertes par l'information seule et décrivant les propriétés des structures d'information optimales dans divers contextes, la conception de structures d'information reste soumise à de nombreuses contraintes pratiques, qui ne sont encore que partiellement prises en compte par la théorie. Comme le soulignent Kamenica, Kim, and Zapechelnyuk (2021) dans leur récent éditorial :

> "La théorie de base [de la conception de structures d'information] repose sur plusieurs hypothèses, qui sont suffisamment plausibles dans divers contextes et ont conduit à de nombreuses découvertes innovantes. Toutefois, nous estimons qu'une approche plus flexible, qui assouplirait

[10]Cette condition, appelée *plausibilité bayésienne* par Kamenica and Gentzkow, est suffisante lorsque le concepteur vise à influencer le comportement d'un seul agent. Lorsque plusieurs agents interagissent stratégiquement après la divulgation de l'information, l'ensemble des croyances pouvant être induites par le concepteur doit également être compatible avec les croyances d'ordre supérieur des joueurs (Mathevet et al., 2020).

[11]Il est important de noter que cette conclusion repose entièrement sur l'hypothèse de l'engagement. Lorsque l'émetteur n'a pas de pouvoir d'engagement (Crawford and Sobel, 1982), sa stratégie de divulgation doit être compatible avec ses *contraintes d'incitation* en plus de la condition de mise à jour bayésienne. Cela implique que la signification des signaux générés par la stratégie de l'expéditeur est déterminée *à l'équilibre*, contrairement au cas de l'engagement où la signification des signaux est déterminée de manière *objective*.

> ces hypothèses, renforcerait considérablement l'applicabilité de la théorie. Nous nous intéressons ici à deux de ces hypothèses. Premièrement, les récepteurs sont des acteurs rationnels standard qui maximisent leur utilité espérée et *effectuent des inférences bayésiennes*. Deuxièmement, il y a *peu ou pas de contraintes sur les structures d'information réalisables*[12]."

Les deux premiers chapitres de cette thèse visent précisément à intégrer à la théorie un nouveau type de contrainte encore inexploré par la littérature, ainsi qu'à assouplir l'hypothèse de mise à jour bayésienne des croyances du récepteur.

Conception de l'information et incitations à l'investissement en amont. Les agents peuvent modifier leur comportement en réaction à la manière dont l'information est structurée *avant même que celle-ci ne soit révélée*. Ces comportements stratégiques *en amont* de la conception de l'information limitent les structures d'information utilisables par le concepteur, car elles doivent être *compatibles avec les incitations des joueurs en amont de la production d'information*. Par exemple, un grand nombre de biens et services sont soumis à des *tests*, mis en place par des régulateurs publics, et destinés à déterminer s'ils peuvent être homologués ou s'ils peuvent bénéficier d'une certification (tels que des labels qualité ou des labels verts). Or, la façon dont les tests sont conçus a un impact en amont sur les incitations des firmes à participer auxdits tests (Rosar, 2017; Harbaugh and Rasmusen, 2018), ou encore sur les comportements de fraude et de falsification (Perez-Richet and Skreta, 2022a,b). Il est donc crucial pour les régulateurs de prendre en compte ces inciations lors de la conception même des tests. Dans le chapitre 1, co-écrit avec Eduardo Perez-Richet, nous examinons un autre type de comportement stratégique en amont : les investissements productifs[13]. Nous élaborons un modèle de conception d'information dont la particularité est que l'état du monde peut être modifié en amont par un agent à un certain coût. L'objectif du concepteur est d'agir uniquement si l'état du monde *final* est suffisamment élevé tandis que celui de l'agent est que le concepteur agisse. Ce modèle s'applique particulièrement aux situations de *sélection*. Par exemple, dans le cas des tests, l'état du monde correspond à la qualité du produit qu'une firme cherche à faire certifier et le concepteur est un régulateur devant décider d'approuver ou non le produit. La structure d'information correspond alors à un test révélant des informations

[12]Traduction de l'auteur.

[13]Bien que ce chapitre soit formulé en termes de problème d'allocation de ressources, nous établissons un résultat d'équivalence entre ce problème de *conception de mécanisme d'allocation* et de *conception de structure d'information* dans la section 1.5.3.

au régulateur sur la qualité du produit après investissement de l'entreprise. Le but du régulateur est d'acquérir la meilleure information possible pour prendre sa décision, tandis que l'entreprise cherche à maximiser la probabilité que son produit soit approuvé par le régulateur. Notre résultat principal montre que les structures d'information optimales pour le concepteur sont *binaires* et *déterministes*. Dans le cas du régulateur, cela correspond à fixer un *seuil* de qualité au-delà duquel un signal signifiant d'approuver le produit est généré par le test et en dessous duquel un signal signifiant de le rejeter est généré. Ce résultat établit une fondation théorique pour les règles de sélection de type "réussite ou échec", largement utilisées dans la pratique, et montre de manière surprenante qu'il est sous-optimal d'introduire de l'aléa dans la structure d'information lorsque des investissements stratégiques peuvent être réalisés en amont de la production d'information, contrairement aux cas de la participation et de la falsification.

Conception de l'information et formation motivée des croyances. Abstraction faite des comportements stratégiques en amont, le comportement du récepteur en aval, lors de la réception de l'information, peut être influencé par la manière dont il *choisit d'interpréter* l'information. Bien que les agents apprennent de l'information qu'ils reçoivent, une vaste littérature en économie comportementale et expérimentale montre qu'ils ne le font pas systématiquement à la manière d'un statisticien, comme le prescrirait la règle de Bayes, mais laissent leurs préférences dicter dans une certaine mesure la façon dont ils forment leurs croyances (voir notamment, Bénabou and Tirole, 2016, p. 150, Epley and Gilovich, 2016, ou Benjamin, 2019, section 9). Nous montrons que ce type de raisonnements *motivés* (selon le terme de Kunda, 1990) est susceptible d'affecter l'efficacité de l'information en tant qu'instrument d'incitation ainsi que la manière dont l'information devrait être révélée par rapport au cas bayésien. Nous basons notre modèle de croyances motivées sur celui proposé par Caplin and Leahy (2019) : après avoir observé un signal informatif, le récepteur forme ses croyances en comparant leur valeur anticipatoire et leur coût psychologique. Cette hypothèse de modélisation reflète d'une part que les croyances ont une *valeur intrinsèque* pour l'agent. Par exemple, il a été montré que les individus associent une utilité aux croyances concernant leur propre image ou leurs perspectives d'avenir, comme se penser meilleur que les autres, être en bonne santé, progresser dans l'échelle sociale, etc. (voir Bénabou and Tirole, 2016). D'autre part, distordre ses croyances par rapport aux croyances bayésiennes est *psychologiquement coûteux*. L'évidence expérimentale montre en effet que les individus emploient des stratégies mentales

sophistiquées et coûteuses pour protéger ou atteindre des croyances souhaitables, comme manipuler leur propre mémoire (voir Bénabou, 2015; Bénabou and Tirole, 2016; Hagenbach and Koessler, 2022, et les références citées par les auteurs) ou refuser de s'exposer à des informations (Golman, Hagmann, and Loewenstein, 2017), telles que des tests fiables permettant de détecter des maladies graves (Oster, Shoulson, and Dorsey, 2013). Dans notre modèle, les croyances du destinataire dépendent donc de ses préférences à la suite de la mise à jour motivée des croyances, et surpondèrent les états du monde associés au gain potentiel le plus élevé. Nous montrons que la distorsion des croyances du récepteur entraîne une distorsion de son comportement. En comparaison d'un individu bayésien, il privilégie les actions associées au gain le plus élevé et à la variabilité de gain la plus élevée, et si une action induit le gain le plus élevé mais qu'une autre induit la plus grande variabilité de gain, la préférence pour l'une ou l'autre dépend de l'ampleur du coût de distortion des croyances du récepteur. Ainsi, *l'efficacité de l'information en tant qu'outil d'incitation peut varier en fonction des enjeux matériels des individus* : la persuasion est plus efficace, par rapport au cas bayésien, lorsqu'elle vise à encourager un comportement risqué mais potentiellement très rentable, et moins efficace lorsqu'elle vise à encourager un comportement plus prudent. Nous illustrons ce résultat avec des applications, montrant pourquoi les campagnes d'information se révèlent souvent inefficaces pour stimuler un investissement accru dans les traitements de santé préventifs et comment les conseillers financiers peuvent bénéficier des croyances excessivement optimistes de leurs clients. Nous montrons également que la divulgation stratégique d'information à des électeurs ayant des préférences partisanes hétérogènes peut entraîner une polarisation des croyances au sein de l'électorat.

Les effets distributionnels de la conception de l'information. le chapitre 3 aborde une question d'une nature différente de celle des deux premiers chapitres. La littérature s'est principalement concentrée sur la caractérisation des situations atteignables par la conception d'information d'un point de vue *agrégé*. Un aspect qui, au contraire, a été négligé concerne les effets *distributionnels* de la conception de l'information, c'est-à-dire comment la révélation d'information affecte différents types d'agents[14]. Nous étudions cette question dans le contexte spécifique des marchés monopolistiques, où la répartition des gains à l'échange générés par l'information est une question de première importance. En effet, les consommateurs laissent constamment des traces de leur identité sur Internet, que ce soit par leur

[14]Une première incursion de la littérature dans cette direction est la contribution de Doval and Smolin (2021).

activité sur les réseaux sociaux, leur utilisation des moteurs de recherche ou leurs achats en ligne. Les données générées et collectées sur les consommateurs sont devenues très précieuses, car elles permettent aux entreprises de *segmenter* les consommateurs, c'est-à-dire de les regrouper en sous-groupes sur la base de caractéristiques observables et de pratiquer une discrimination par les prix en fonction du sous-groupe. Comme mentionné précédemment dans l'introduction, Bergemann et al. (2015) montrent que la révélation d'information sur la disposition marginale à payer des consommateurs peut induire une grande variété de résultats en termes de ségementation et de surplus social. En particulier, il est toujours possible de segmenter la population des consommateurs de façon à leur allouer l'entièreté des gains à l'échange. Pour ce faire, il est possible de créer des segments qui regroupent les consommateurs dont la disposition marginale à payer est élevée et ceux dont la disposition marginale à payer est faible. Une telle segmentation force le vendeur à pratiquer des prix plus bas sur ces segments[15] et, par conséquent, à faire bénéficier les consommateurs ayant une forte disposition marginale à payer de prix plus bas. Toutefois, un inconvénient majeur de ce type de segmentation est que si la disposition marginale à payer et la richesse des consommateurs sont positivement corrélées, les segmentations maximisant le surplus des consommateur ont tendance à *profiter aux consommateurs les plus riches*. Comment concevoir l'information du vendeur de manière à bénéficier aux consommateurs les plus pauvres ? Nous répondons à cette question en étudiant un modèle de segmentation d'un marché monopolistique avec un objectif redistributif. Nous montrons que les segmentations redistributives optimales génèrent toujours des allocations efficientes au sens de Pareto, mais peuvent nécessiter d'accorder une part strictement positive du surplus au vendeur. Nous identifions l'ensemble des marchés pour lesquels la segmentation redistributive implique de laisser une rente au monopole. Nous élaborons également une procédure permettant de construire la segmentation redistributive optimale et montrons que, lorsque la corrélation entre la disposition à payer et la richesse des consommateurs et suffisamment élevée, la segmentation redistributive optimale divise les consommateurs en segments contigus : les consommateurs sont classés et regroupés de manière croissante en fonction de leur disposition à payer. De cette façon, les consommateurs plus pauvres bénéficient de plus faibles prix que les plus riches. De manière intéressante, ce résultat contraste fortement avec les segmentations maximisant le surplus des consommateurs sans préoccupation distributionnelle, qui regroupent ensemble consommateurs riches en pauvres afin

[15]Sinon, le monopole diminuerait trop la marge extensive par rapport à la marge intensive, en excluant trop d'acheteurs de la consommation, ce qui diminuerait ses profits.

de faire bénéficier les plus riches de prix plus faibles.

Bibliographie

Ball, I. (2019) : “Scoring Strategic Agents,” *arXiv*.

Bénabou, R. (2015) : “The Economics of Motivated Beliefs,” *Revue d’économie politique*, 125, 665–685.

Bénabou, R. and J. Tirole (2016) : “Mindful Economics : The Production, Consumption, and Value of Beliefs,” *Journal of Economic Perspectives*, 30, 141–164.

Benjamin, D. J. (2019) : “Errors in probabilistic reasoning and judgment biases,” in *Handbook of Behavioral Economics : Applications and Foundations 2*, Elsevier B.V., vol. 2, chap. 2, 69–186.

Bergemann, D., B. Brooks, and S. Morris (2015) : “The Limits of Price Discrimination,” *American Economic Review*, 105, 921–57.

Bergemann, D. and S. Morris (2016) : “Bayes correlated equilibrium and the comparison of information structures in games,” *Theoretical Economics*, 11, 487–522.

——— (2019) : “Information Design : A Unified Perspective,” *Journal of Economic Literature*, 57, 44–95.

Best, J. and D. Quigley (2017) : “Persuasion for the Long Run,” *SSRN Electronic Journal*.

Boleslavsky, R. and C. Cotton (2015) : “Grading Standards and Education Quality,” *American Economic Journal : Microeconomics*, 7, 248–279.

Caplin, A. and J. Leahy (2019) : “Wishful Thinking,” *NBER Working Paper Series*.

Che, Y.-K. and J. Hörner (2018) : “Recommender Systems as Mechanisms for Social Learning,” *Quarterly Journal of Economics*, 133, 871–925.

Crawford, V. P. and J. Sobel (1982) : “Strategic Information Transmission,” *Econometrica*, 50, 1431–1451.

Doval, L. and A. Smolin (2021) : “Information Payoffs : An Interim Perspective,” *Working paper*.

ELY, J. C. (2017) : “Beeps,” *American Economic Review*, 107, 31–53.

EPLEY, N. AND T. GILOVICH (2016) : “The Mechanics of Motivated Reasoning,” *Journal of Economic Perspectives*, 30, 133–140.

GENTZKOW, M. AND E. KAMENICA (2014) : “Costly Persuasion,” *American Economic Review : Papers & Proceedings*, 104, 457–462.

——— (2017) : “Competition in Persuasion,” *Review of Economic Studies*, 84, 300–322.

GEORGIADIS, G. AND B. SZENTES (2020) : “Optimal Monitoring Design,” *Econometrica*, 88, 2075–2107.

GOLDSTEIN, I. AND Y. LEITNER (2018) : “Stress tests and information disclosure,” *Journal of Economic Theory*, 177, 34–69.

GOLMAN, R., D. HAGMANN, AND G. LOEWENSTEIN (2017) : “Information Avoidance,” *Journal of Economic Literature*, 55, 96–135.

GROSSMAN, S. J. (1981) : “The Informational Role of Warranties and Private Disclosure about Product Quality,” *Journal of Law and Economics*, 24, 461–483.

HAGENBACH, J. AND F. KOESSLER (2022) : “Selective memory of a psychological agent,” *European Economic Review*, 142, 104012.

HARBAUGH, R. AND E. RASMUSEN (2018) : “Coarse Grades : Informing the Public by Withholding Information,” *American Economic Journal : Microeconomics*, 10, 210–235.

HERNÁNDEZ, P. AND Z. NEEMAN (2022) : “How Bayesian Persuasion Can Help Reduce Illegal Parking and Other Socially Undesirable Behavior,” *American Economic Journal : Microeconomics*, 14, 186–215.

ICHIHASHI, S. (2023) : “Privacy, Transparency, and Policing,” *Working paper*.

INOSTROZA, N. (2019) : “Persuading Multiple Audiences : An Information Design Approach to Banking Regulation,” *SSRN Electronic Journal*.

KAMENICA, E. (2019) : “Bayesian Persuasion and Information Design,” *Annual Review of Economics*, 11, 249–272.

KAMENICA, E. AND M. GENTZKOW (2011) : “Bayesian Persuasion,” *American Economic Review*, 101, 2590–2615.

Kamenica, E., K. Kim, and A. Zapechelnyuk (2021) : "Bayesian persuasion and information design : perspectives and open issues," *Economic Theory*, 72, 701–704.

Kunda, Z. (1990) : "The case for motivated reasoning." *Psychological Bulletin*, 108, 480–498.

Lin, X. and C. Liu (2022) : "Credible Persuasion," in *Proceedings of the 23rd ACM Conference on Economics and Computation*, New York, NY, USA : ACM, 469–469.

Lipnowski, E., D. Ravid, and D. Shishkin (2022) : "Persuasion via Weak Institutions," *Journal of Political Economy*, 130, 2705–2730.

Mathevet, L., D. Pearce, and E. Stacchetti (2022) : "Reputation for a degree of honesty," *Working paper*.

Mathevet, L., J. Perego, and I. Taneva (2020) : "On Information Design in Games," *Journal of Political Economy*, 128, 1370–1404.

Milgrom, P. R. (1981) : "Good News and Bad News : Representation Theorems and Applications," *Bell Journal of Economics*, 12, 380.

Orlov, D., P. Zryumov, and A. Skrzypacz (2021) : "The Design of Macro-Prudential Stress Tests," *SSRN Electronic Journal*.

Oster, E., I. Shoulson, and E. R. Dorsey (2013) : "Optimal Expectations and Limited Medical Testing : Evidence from Huntington Disease," *American Economic Review*, 103, 804–830.

Ostrizek, F. (2022) : "Vague by Design : Performance Evaluation and Learning from Wages," *Working paper*.

Perez-Richet, E. and V. Skreta (2022a) : "Fraud-proof non-market allocation mechanisms," *Working paper*.

——— (2022b) : "Test Design Under Falsification," *Econometrica*, 90, 1109–1142.

Romanyuk, G. and A. Smolin (2019) : "Cream Skimming and Information Design in Matching Markets," *American Economic Journal : Microeconomics*, 11, 250–276.

Rosar, F. (2017) : "Test design under voluntary participation," *Games and Economic Behavior*, 104, 632–655.

Saeedi, M. and A. Shourideh (2020) : "Optimal Rating Design," *arXiv*.

Spence, M. (1973) : "Job Market Signaling," *Quarterly Journal of Economics*, 87, 355–374.

Vellodi, N. (2018) : "Ratings Design and Barriers to Entry," *SSRN Electronic Journal*.

Zapechelnyuk, A. (2020) : "Optimal Quality Certification," *American Economic Review : Insights*, 2, 161–176.

1. Non-Market Allocation Mechanisms: Optimal Design and Investment Incentives[1]

Abstract

We study how to optimally design selection mechanisms, accounting for agents' investment incentives. A principal wishes to allocate a resource of homogeneous quality to a heterogeneous population of agents. The principal commits to a possibly random selection rule that depends on a one-dimensional characteristic of the agents she intrinsically values. Agents have a strict preference for being selected by the principal and may undertake a costly investment to improve their characteristic before it is revealed to the principal. We show that even if random selection rules foster agents' investments, especially at the top of the characteristic distribution, deterministic "pass-fail" selection rules are in fact optimal.

1.1. Introduction

Allocating goods, services, or prizes often requires selecting agents on the basis of measurable characteristics. In many important contexts, including university admissions, the allocation of research grants, the issuance of certifications by regulatory agencies, and promotion decisions within organizations, selection procedures inherently generate incentives for agents to strategically *invest* in the characteristics on which they are evaluated. When such investments are

[1]This chapter is a joint work with Eduardo Perez-Richet. We thank Ricardo Alonso, Daniel Barreto, Aislinn Bohren, Alexis Ghersengorin, Simon Gleyze, Jeanne Hagenbach, Emeric Henry, Emir Kamenica, Jan Knoepfle, Frédéric Koessler, Flavien Léger, Shengwu Li, Laurent Mathevet, Daniel Monte, Franz Ostrizek and Sevgi Yuksel for their valuable comments and suggestions at various stages of the project. We also thank all the seminar participants at Sciences Po, Paris School of Economics, European University Institute, CSEF – University of Naples Federico II, and Institute for Microeconomics at the University of Bonn. Part of this research was conducted while both authors were visiting the Department of Economics at the European University Institute, whose hospitality is gratefully acknowledged. All remaining errors are ours. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 850996 – MOREV and 101001694 – IMEDMC).

possible, institutions face the challenge of designing selection mechanisms that take into account the *endogenous* nature of agents' characteristics. Accounting for *investment incentives* is therefore of paramount importance in designing effective selection mechanisms. We propose a theory in which agents can transform their characteristics at some cost in response to selection and answer the research question: What is the optimal selection mechanism for an institution whose objective is to select agents with the highest possible characteristic, taking into account agents' investment incentives?

A simple selection rule, often used in practice, is to set a *pass-fail* selection cutoff, such as exams with a pass grade or certifications with a fixed quality standard. Pass-fail selection induces agents just below the cutoff to invest so they can pass but also discourages investments for all agents above the cutoff. Intuitively, random selection rules could perform better by spreading investment incentives, especially at the top. Contrary to this intuition, our main result shows that, in fact, pass-fail selection is *optimal* when accounting for investment. By doing so, we provide a firmer foundation for the use of this simple class of selection rules.

We consider the following model. A principal wishes to allocate a unit mass of resources to a unit mass of agents. To do so, she commits to a possibly random selection rule based on a one-dimensional characteristic of agents. However, she cannot use monetary transfers. The principal intrinsically values the characteristic of agents so long as it exceeds a given preference threshold. Agents have a strict preference for being allocated the good and can undertake a costly investment to improve their characteristic before it is revealed. Finally, the selection rule operates and the outcome is realized.

We make two assumptions. First, the principal's preference threshold lies in the upper tail of the distribution of characteristics, so that its density decreases in the region of interest. For tractability, we also assume that the investment cost function of the agents is quadratic in the amount of investment, and that the value of allocating the good for the principal is linear in the characteristic.

To give an intuition for our result, consider deviating from a pass-fail rule with a given cutoff to an increasing random rule allocating the resource with non-zero probability above the cutoff. Randomizing the allocation has a positive effect on the principal's payoff by encouraging investments at the top of the characteristic distribution. However, it has a negative effect by unduly rejecting some agents having invested, and by excluding agents at the bottom of the distribution who would have invested under the pass-fail rule. When the distribution of initial characteristics has a decreasing density, the negative effect always dominates.

Technically, the non-linearity of the principal's objective created by the absence of monetary transfers makes the characterization of optimal mechanisms difficult in our setting. In contrast to environments with transfers and quasi-linear utilities, we cannot use the standard resolution method developed in Myerson (1981). To establish the optimality of pass-fail selection, we start by showing that any optimal selection rule must be zero below the principal's preference threshold and non-decreasing above. We then consider a transformation of the agents' indirect utility function that we call pseudo-utility. We characterize the set of pseudo-utility functions that are implementable under incentive-compatible mechanisms. Using this characterization we then show that the original optimization program of the principal boils down to a problem of calculus of variations with pseudo-utility as an optimization variable. This problem is not standard because it involves maximizing a convex objective functional, and thus cannot be solved using the first-order approach. We prove that the set of implementable pseudo-utility functions is compact and convex. Hence, the Krein-Milman theorem and Bauer's Maximum Principle guarantee that a solution to the variational program can be found at an extreme point of the domain. We provide necessary conditions on the shape of extreme points. Together with the tangent inequality for convex functionals, these necessary conditions imply that the extreme point corresponding to the pseudo utility implemented by a pass-fail selection rule is an optimal solution. The final step simply requires optimizing the principal's payoff with respect to the one-dimensional selection cutoff.

Next, we perform a comparative statics exercise with respect to the magnitude of agents' investment costs. We show that the optimal cutoff decreases and that the mass of excluded agents at the bottom increases as investment costs increase. Accordingly, the designer's equilibrium payoff decreases in the agents' costs and naturally converges to the optimal payoff she would obtain if she could not induce any investment as costs become arbitrarily large. Under a pass-fail rule, a strictly positive mass of agents are bunching at the selection cutoff. We prove that agents with lower types in the bunching interval are affected negatively by an increase in investment costs while agents with higher types in the interval are affected positively. Intuitively, the direct effect of an upward scaling in the costs is greater than the indirect effect of the decrease in the selection cutoff for all types which are already sufficiently close to it and conversely for lower types.

We then extend the model in three directions. First, we consider the case where the principal is subject to a capacity constraint, i.e., has a strictly smaller amount of resources to allocate than the total mass of agents. Second, we consider the

problem of a planner maximizing utilitarian social welfare. We show that in both cases, pass-fail mechanisms remain optimal. In the first case, the optimal cutoff is naturally higher than in the baseline solution whenever the capacity constraint is binding. In the case of optimal utilitarian welfare, however, accounting for the agents' investment costs pushes the optimal cutoff downwards. Finally, we show that the optimal outcome can still be implemented when the principal's commitment power is relaxed. We assume that the principal cannot commit to allocation mechanisms anymore. Instead, she makes her allocation decision based on information provided by an intermediary. The intermediary shares the same objective as the principal and can produce information on the results of the agents' investments by committing to a statistical experiment (Blackwell, 1951, 1953). After observing the experiment chosen by the intermediary, the agents choose their investment strategies. We show that the recommendation principle holds in this environment. This implies that the intermediary can restrict her choice, without loss of generality, to statistical experiments whose outcomes are obedient action recommendations for the principal. Because of the alignment of the preferences between the intermediary and the principal, the obedience constraint is never binding. The conditional probability of recommending to allocate under the intermediary's policy can therefore be interpreted as the selection rule in our original problem. This exercise shows that committing to information or to a mechanism is equivalent in this environment. Consequently, a higher commitment power has no additional value for the principal.

Relation to the literature. The question of investment incentives in resource allocation mechanisms has been the subject of an extensive literature, particularly in the context of the "hold-up" problem.[2] A fundamental contribution is the one of Rogerson (1992) who shows that VCG allocation mechanisms (Vickrey, 1961; Clarke, 1971; Groves, 1973) induce ex-ante optimal investment incentives and thus overcome the hold-up problem. This result has been extended by Bergemann and Välimäki (2002) to situations where agents invest in information acquisition before participating in the mechanism, by Athey and Segal (2013) to dynamic environments, by Hatfield, Kojima, and Kominers (2019) and Akbarpour, Kominers,

[2]The hold-up problem arises in situations where (i) the parties to a future transaction can undertake specific sunk cost investments that affect the value of the transaction and (ii) the form and value of the optimal transaction may be affected by unforeseen and non-contractible contingencies. Historically, the hold-up problem originates from the literature on transaction costs and the nature of the firm (Klein, Crawford, and Alchian, 1978; Williamson, 1979) and has received a lot of attention from the literature on incomplete contracts (e.g., Grossman and Hart, 1986; Tirole, 1986; Hart and Moore, 1988; Chung, 1991; MacLeod and Malcomson, 1993; Aghion, Dewatripont, and Rey, 1994).

Li, Li, and Milgrom (2022) to approximately efficient mechanisms, and by Tomoeda (2019) to full implementation. Investment incentives have also been studied in more specific settings such as public procurement (Laffont and Tirole, 1986; Arozamena and Cantillon, 2004), revenue-maximizing auctions (Daley, Schwarz, and Sonin, 2012; Gershkov, Moldovanu, Strack, and Zhang, 2021), bilateral trading (Gul, 2001; Lau, 2008; Dilmé, 2019; Condorelli and Szentes, 2020) and matching mechanisms (Hatfield, Kojima, and Kominers, 2014; Hatfield, Kojima, and Narita, 2016). For the most part, the aforementioned works consider allocation problems with transfers, and all consider investments as a costly action influencing the agents' valuations (or costs) before participating in the mechanism. In contrast, we consider mechanisms without transfers where the agents' investments do not affect their valuations for the object but have an intrinsic value for the designer.

We thus also contribute to the literature analyzing the optimal design of resource allocation mechanisms without monetary transfers, often referred to as "non-market mechanisms".[3] A significant part of the literature focuses on the costly signaling setting, in which the designer can require agents to engage in a socially wasteful activity (e.g., waiting in line or filling application forms) used as a screening device to separate high types from low types, such as in Hartline and Roughgarden (2008), Yoon (2011), Condorelli (2012), Chakravarty and Kaplan (2013), Ashlagi, Monachou, and Nikzad (2021a,b), Ottaviani (2021) or Kleiner et al. (2021), Section 4.1.[4] Relatedly, Perez-Richet and Skreta (2022b) show that selection rules inducing falsification are optimal when agents can misreport their types at some cost to the designer. Perez-Richet and Skreta (2022a), in contrast, focus on the design of falsification-proof selection rules. Other studies highlight the fact that correlation can be profitably exploited when transfers are not available. In Bhaskar and Sadler (2020) the designer takes advantage of the fact that allocating certain types of goods causes positive externalities, thus correlating agents' valuations. Similarly, in Kattwinkel (2020), Kattwinkel, Niemeyer, Preusser, and Winter (2022) and Niemeyer and Preusser (2022) the optimal mechanisms leverage on the correlation

[3]It is worth noting that a parallel literature has focused on establishing the optimality of non-market mechanisms when a trade-off between allocative efficiency and equity is involved. A seminal contribution in that strand of the literature is Weitzman (1977). Condorelli (2013) shows that non-market mechanisms are optimal when the characteristic valued by the principal is not sufficiently correlated with the agents' willingness to pay, preventing her from obtaining all relevant information through the appropriate design of prices. Akbarpour, Dworczak, and Kominers (2020) extend Condorelli's contribution to environments with a continuum of heterogeneous qualities, a continuum of agents, and endogenous Pareto weights reflecting the statistical correlation between the agents' willingness to pay and their marginal contribution to social welfare. Kleiner, Moldovanu, and Strack (2021) also extend it to general matching contests.

[4]Ambrus and Egorov (2017) and Amador and Bagwell (2020) also characterize optimal delegation mechanisms with signaling.

of agents' information. When information cannot be extracted through prices, the mechanism designer can also rely on costly verification such as in Ben-Porath, Dekel, and Lipman (2014), Mylovanov and Zapechelnyuk (2017), Erlanson and Kleiner (2019), and Kattwinkel and Knoepfle (2022).[5] In contrast to all these papers, we analyze a model in which the designer selects agents on the basis of the observable outcome of an investment. This constitutes the polar case to that of signaling or falsification. Indeed, our model can be viewed as a situation in which the agent's costly action does not serves as a pure screening variable but is instead intrinsically valuable.

In our setting, the optimal selection rule is deterministic. This is in stark contrast with optimal mechanisms under costly signaling, which generally involve random rationing because of binding incentive constraints (see Hartline and Roughgarden, 2008; Yoon, 2011; Condorelli, 2012; Chakravarty and Kaplan, 2013; Ashlagi et al., 2021a,b; Kleiner et al., 2021). Similarly, falsification incentives induce the designer to use random selection rules in Perez-Richet and Skreta (2022a,b). Under correlated information, Kattwinkel (2020) and Niemeyer and Preusser (2022) show that optimal mechanisms might involve randomization and may not even be monotone. Threshold selection rules, however, turn out to prove optimal when the designer can verify the agents' claims at some cost as in Ben-Porath et al. (2014), Mylovanov and Zapechelnyuk (2017), Erlanson and Kleiner (2019) and Kattwinkel and Knoepfle (2022).[6] In a setting with a continuum of heterogeneous qualities and a continuum of agents with different preference intensities, Ortoleva, Safonov, and Yariv (2021) show that random selection rules are optimal under both symmetric and asymmetric information about the agents' preferences. Our result also resonates with the optimality of posted price mechanisms in settings with monetary transfers, proved independently by Myerson (1981) for revenue-maximizing auctions and by Riley and Zeckhauser (1983) in the context of monopoly pricing.[7]

Interestingly, our paper also relates to the literature on statistical discrimination and affirmative action.[8] Our problem of allocation without transfers can be seen as a generalization of Chan and Eyster (2003) and Ray and Sethi (2010) where the

[5]Halac and Yared (2020) conduct a similar analysis in the context of delegation.

[6]Threshold mechanisms also turn out to be optimal in the model of delegation with costly verification of Halac and Yared (2020). We also refer to Kováč and Mylovanov (2009) and Kleiner et al. (2021), Section 4.2, who prove under which conditions optimal delegation involves randomization.

[7]Börgers, Krähmer, and Strausz (2015) give an alternative and elegant proof of that result by showing that posted price mechanisms correspond to extreme points of the space of allocation functions.

[8]We refer to Fang and Moro (2011) for an in-depth review of that literature, and to Onuchic (2022) for a review of recent theoretical contributions.

distribution of test scores is endogenous. We thus provide a microfoundation for Chan and Eyster's restriction to monotone admission rules based on the agents' investment incentive-compatibility. The selection mechanisms in the models of Chan and Eyster and Ray and Sethi are also deterministic. However, both papers recognize the possibility that color-blind affirmative action constraints can make the optimal rule non-monotone. To the best of our knowledge, Fryer, Loury, and Yuret (2007) and Fryer and Loury (2013) are the only papers to study models of optimal selection where applicants can undertake investments that really affect their types. Our model however, considers a more general form of investment technology with a continuum of skills. It is interesting to note that the optimal selection rule is also pass-fail in their settings.

Our work also echoes an emerging literature at the intersection of economics and computer science, studying how to optimally design linear classifiers when the input features are manipulable by an agent. Hu, Immorlica, and Vaughan (2018), Ball (2019) and Frankel and Kartik (2022) study the optimal design of linear selection rules when the agent has the ability to falsify its type at some cost. More closely connected to our paper, Kleinberg and Raghavan (2020) study how to design a linear classifier so as to induce the agents to invest some effort to improve their outcomes as opposed to gaming the classifier.

On the methodological front, we solve a problem of calculus of variations which shares a similar structure to convexity-constrained variational problems arising in monopolistic screening models (see, e.g., Rochet and Choné, 1998; Carlier, 2001; Manelli and Vincent, 2007; Daskalakis, Deckelbaum, and Tzamos, 2017; Kleiner and Manelli, 2019; Bergemann and Strack, 2022). However, instead of being linear as in the previously mentioned papers, the objective functional of our variational program is convex. We follow a similar approach to Manelli and Vincent (2007) and Kleiner et al. (2021) by characterizing an optimal solution among the extreme points of a compact convex functional space.

1.2. Model

A principal (designer, she) has unit mass of resources to allocate to a unit mass of agents.[9] The agents can undertake investments resulting in a new type that is observed by the designer. The designer commits ex-ante to an selection rule contingent on the agents' final types. Her value from selecting an agent is his final type, and her outside option is to not allocate the good at all in which case she gets

[9]We show in section 1.5 that our main result still holds when the designer is capacity constrained.

a null payoff. The agents only care about getting the good independently of their types and their investment cost is increasing and convex in the type improvement.

Types. There is a continuum of agents characterized by a *type* $\theta \in \Theta = [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}$. We set $\underline{\theta} < 0 < \bar{\theta}$. The total mass of agents is normalized to one. Types are distributed according to the cumulative distribution function $F \colon \Theta \to [0, 1]$. We assume that F admits a density function $f \colon \Theta \to \mathbb{R}$ which is strictly positive and continuously differentiable on the support Θ.

Investments. The agents can transform their types at some cost. Acquiring a final type $t \in T = \mathbb{R}$ entails a cost $\gamma c(t, \theta)$ to an agent with initial type θ, where $\gamma \in \mathbb{R}_+$ is a scaling parameter, and:

$$c(t, \theta) = \frac{\max\{0, (t - \theta)\}^2}{2},$$

for all $(t, \theta) \in T \times \Theta$. Under this specification for the cost function, the transformed type can indeed be regarded as the outcome of an *investment*. The cost for agents to acquire a new type is non-negative only if it is higher than their initial type, and is increasing and convex as a function of the type increase. Moreover, the cost exhibits decreasing differences, i.e., $c(t', \theta') - c(t, \theta') \leq c(t', \theta) - c(t, \theta)$ for any $t' > t$ and $\theta' > \theta$, with strict inequality so long as $t > \theta$.[10] This property of the cost function captures heterogeneity in the investment ability of agents. It is marginally costlier for an agent whose initial type is low to acquire a high final type than for an agent whose initial type is already high.

Payoffs. The designer can choose to allocate or not the resource to each agent. Her allocation choice is denoted $a \in A = \{0, 1\}$ and her payoff function is given by $v(a, t) = at$ for any $(a, t) \in A \times T$. That is, the designer's payoff from allocating the resource to an agent with final type t is normalized to t while her payoff from not allocating the resource is set to zero.[11] All agents have the same preference over allocations. They receive a payoff normalized to one upon allocation, and get zero otherwise. The payoff of an agent with initial type θ is thus given by the allocation choice of the designer net of the investment cost $u(a, t, \theta) = a - \gamma c(t, \theta)$.

[10] Such a specification for the cost function is also present in the signaling models of Frankel and Kartik (2019, 2022) and Ball (2019) but result in a different interpretation. In their papers, the type of agents is multidimensional. The first dimension θ is identified with the agents' "natural" ability while the second dimension γ captures their ability to "game" the designer's selection rule.

[11] Setting the non-allocation payoff to zero is equivalent to saying that the designer has no intrinsic utility for the good.

Mechanisms and incentive-compatibility. The designer cannot observe θ but knows F and can observe the final type t resulting from each agents' investment. She commits ex-ante to an *selection rule* $\sigma : T \to [0, 1]$ specifying the probability of selecting an agent with final type t. After observing the selection rule σ, agents choose an *investment rule* $\tau : \Theta \to T$. Any pair (σ, τ) is called a *mechanism*. We say that an investment rule is *incentive-compatible* if it maximizes the probability of allocation net of investment costs for all initial types.

Definition 1 (Incentive-compatibility)**.** *An investment rule $\tau : \Theta \to T$ is incentive-compatible under the selection rule $\sigma : T \to [0, 1]$ if:*

$$\tau(\theta) \in \arg\max_{t \in T} \sigma(t) - \gamma c(t, \theta), \tag{IC}$$

for all $\theta \in \Theta$.

We say that an investment rule τ is *implementable* if there exists an selection rule σ under which τ is *incentive-compatible*.

Timing. The timing of the game is the following:

(i) **Allocation rule:** The designer commits to an selection rule σ which is publicly observed.

(ii) **Types:** The agents' types are drawn according to the cumulative distribution function F.

(iii) **Investments:** Each agent privately observes its type θ and undertakes an investment taking into account its cost $\gamma c(t, \theta)$ as well as the selection rule σ. The investment results in a new type $\tau(\theta)$.

(iv) **Outcome and payoffs:** The designer observes $\tau(\theta)$. The mechanism (σ, τ) generates an outcome $x(\theta) = \sigma(\tau(\theta))$ specifying all agents' allocation probabilities given their investments, and payoffs are realized.

Design problem. Given a mechanism (σ, τ) and the density function f, the ex-ante expected payoff of the designer is given by the expected final type conditional on allocation:

$$V(\sigma, \tau) = \int_{\underline{\theta}}^{\bar{\theta}} \tau(\theta)\, \sigma(\tau(\theta)) f(\theta)\, \mathrm{d}\theta.$$

The problem of the designer consists in maximizing the expected final type among selected agents, taking into account the agents' investment incentives. Formally, this corresponds to the following optimization program:

$$\underset{\sigma,\tau}{\text{maximize}}\ V(\sigma,\tau) \text{ subject to (IC).} \tag{P}$$

1.3. Main results

We start by distinguishing deterministic from random mechanisms and showing that we can restrict our analysis without loss of generality to the class of monotone selection rules. Pass-fail rules correspond to the class of monotone and deterministic selection rules. We then formulate assumptions on the designer's preferences and the magnitude of the agents' costs, and state our main result. We also explore some properties of the optimal mechanism.

1.3.1. Optimality of pass-fail selection rules

Deterministic vs. random allocations. An selection rule is called *deterministic* if it allocates the resource to a strictly positive measure of types with probability one and excludes others for sure. Conversely, an selection rule is *random* if it is not deterministic, i.e., the resource is allocated with an interior probability for a non-negligible measure of types. We give the formal definition below.

Definition 2. *An selection rule σ is deterministic if $\sigma(t) \in \{0,1\}$ for almost all $t \in T$. An selection rule is random if there exists a (Lebesgue) measurable subset $\tilde{T} \subset T$ with strictly positive measure such that $0 < \sigma(t) < 1$ for all $t \in \tilde{T}$.*

Restriction to monotone allocations. We show in the next lemma that we can restrict our analysis without loss of generality to monotone increasing selection rules assigning zero probability to strictly negative types.

Lemma 1 (Monotone selection rules)**.** *One can, without loss of generality, restrict attention to mechanisms such that:*

(i) $\sigma(t) = 0$ *for all* $t \in\ ]-\infty, 0[$*, and;*

(ii) σ *is non-decreasing on* $[0, +\infty[$.

Any selection rule satisfying properties (i) and (ii) is called monotone.

Proof. See appendix A.2.1 $\square$

The first property is implied by optimality. Indeed, the designer always loses from selecting agents with a type below her preference threshold. Monotonicity is an implication of (IC). Since the investment cost function satisfies decreasing differences, any incentive-compatible investment rule must be increasing. Moreover, the cost function is non-decreasing in the final type. As a result, letting the selection rule σ be decreasing on some interval would necessarily violate (IC).

Pass-fail mechanisms. An selection rule is *pass-fail* if there exists a cutoff $t^\dagger$ above which all agents with a final type greater than $t^\dagger$ obtain the resource with probability one, and all agents whose final types are strictly less than $t^\dagger$ obtain the resource with probability zero. Here is the formal definition.

Definition 3. *Let $t^\dagger \in T$. An selection rule σ is a $t^\dagger$-*pass-fail *rule if:*

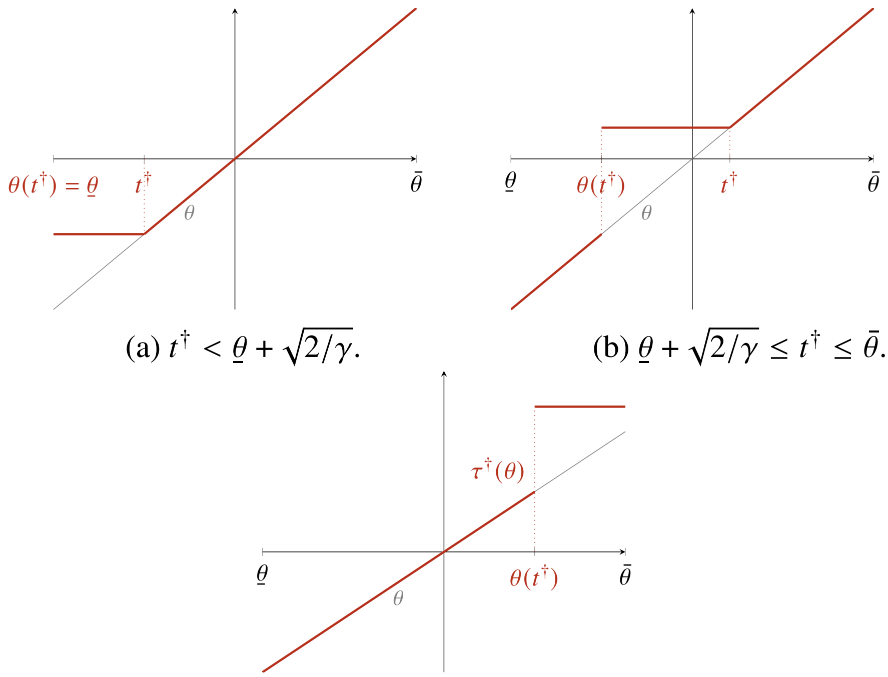
$$\sigma(t) = \mathbb{1}\{t \geq t^\dagger\},$$

for any $t \in T$.

Let σ be a pass-fail rule with allocation cutoff $t^\dagger$. Observe that varying the cutoff $t^\dagger$ from zero to infinity (essentially) spans the entire family of *monotone and deterministic* selection rules.

We now show that the investment rule implemented by a pass-fail selection rule is essentially unique. Let $\theta(t^\dagger)$ be the initial type $\theta \in \Theta$ solving the equation $\gamma c(t^\dagger, \theta) = 1$ when it exists. All agents with initial types given by $\theta(t^\dagger)$ are thus indifferent between keeping type $\theta(t^\dagger)$ at zero cost and acquiring the final type $t^\dagger$ at a cost equal to 1. Given the quadratic form of the cost function the threshold $\theta(t^\dagger)$ is equal to $t^\dagger - \sqrt{2/\gamma}$ whenever $\underline{\theta} + \sqrt{2/\gamma} \leq t^\dagger \leq \bar{\theta} + \sqrt{2/\gamma}$ and we set $\theta(t^\dagger) = \underline{\theta}$ whenever $t^\dagger < \underline{\theta} + \sqrt{2/\gamma}$. First, agents whose initial type is below $\theta(t^\dagger)$ have a cost which is too high to acquire the final type $t^\dagger$ and, as a result, are rejected by the designer and keep their initial type at zero cost. Agents with a type in between $\theta(t^\dagger)$ and $t^\dagger$, in turn, all choose to invest in the minimal final type type guaranteeing admission with probability one, which corresponds exactly to $t^\dagger$. Finally, the agents whose initial types are above $t^\dagger$ are approved with probability one at zero cost. Therefore, the following investment rule corresponds to the the (essentially) unique investment rule that is implementable by a $t^\dagger$-pass-fail selection rule:

Therefore, the following investment rule (illustrated on figure 1.1) corresponds to the the (essentially) unique investment rule that is implementable by a $t^\dagger$-pass-fail



(a) $t^\dagger < \underline{\theta} + \sqrt{2/\gamma}$.

(b) $\underline{\theta} + \sqrt{2/\gamma} \leq t^\dagger \leq \bar{\theta}$.

(c) $t^\dagger > \bar{\theta}$

Figure 1.1: Incentive-compatible investment under a $t^\dagger$-pass-fail rule.

selection rule:

$$\tau^\dagger(\theta) = \begin{cases} \theta & \text{if } \theta \in [\underline{\theta}, \theta(t^\dagger)[\\ t^\dagger & \text{if } \theta \in [\theta(t^\dagger), t^\dagger[\\ \theta & \text{if } \theta \in [t^\dagger, \bar{\theta}] \end{cases} \tag{1.1}$$

if $t^\dagger < \bar{\theta}$, and

$$\tau^\dagger(\theta) = \begin{cases} \theta & \text{if } \theta \in [\underline{\theta}, \theta(t^\dagger)[\\ t^\dagger & \text{if } \theta \in [\theta(t^\dagger), \bar{\theta}[\end{cases} \tag{1.2}$$

if $t^\dagger \geq \bar{\theta}$. A $t^\dagger$-pass-fail rule thus segments the population of agents in two categories: the agents who keep their initial types at no cost and those who bunch at the allocation cutoff $t^\dagger$. Any mechanism (σ, τ) such that σ is a $t^\dagger$-pass-fail rule and τ is either given by equation (1.1) if $t^\dagger < \bar{\theta}$ or by equation (1.2) if $t^\dagger \geq \bar{\theta}$ is called a *$t^\dagger$-pass-fail mechanism.*

Assumptions. Let us start by defining $\theta_0 = \min\{\theta \in \Theta \,|\, c(0, \theta) \leq 1\}$. The type θ_0 corresponds to the lowest initial type for which investing at the designer preference threshold would be feasible if the good were allocated with probability one. Given the quadratic form of the cost, we have the closed-form solution $\theta_0 = -\sqrt{2/\gamma}$. Any agent with an initial type strictly lower than θ_0 thus has too

high a cost to reach any type above the designer's preference threshold under any monotone selection rule. Therefore, any agents with an initial type below θ_0 must be rejected with probability one under the optimal allocation. We introduce our first assumption, which concerns the designer's preference threshold.

Assumption 1. *The density function f is non-increasing on the interval $[\theta_0, \bar{\theta}]$.*

This assumption has two implications. First, the density function is decreasing on the right tail of the type distribution. Second, it implies that the designer's preference threshold lies sufficiently to the right of that tail, so that the density function is non-increasing over the interval $[\theta_0, \bar{\theta}]$. We also make an assumption on the magnitude of the agents' investment costs.

Assumption 2. *The scaling parameter γ satisfies $\gamma > 1/c(0, \underline{\theta})$.*

This assumption clarifies the exposition of our results by eliminating cases where the magnitude of the investment costs would be so small that the designer would allocate the resource to all agents under the optimal rule. However, it is not necessary to prove our main result. Formally, this assumption ensures that θ_0 is bounded away from $\underline{\theta}$, so we can exclude all the types in the interval $[\underline{\theta}, \theta_0[$ from our subsequent analysis.

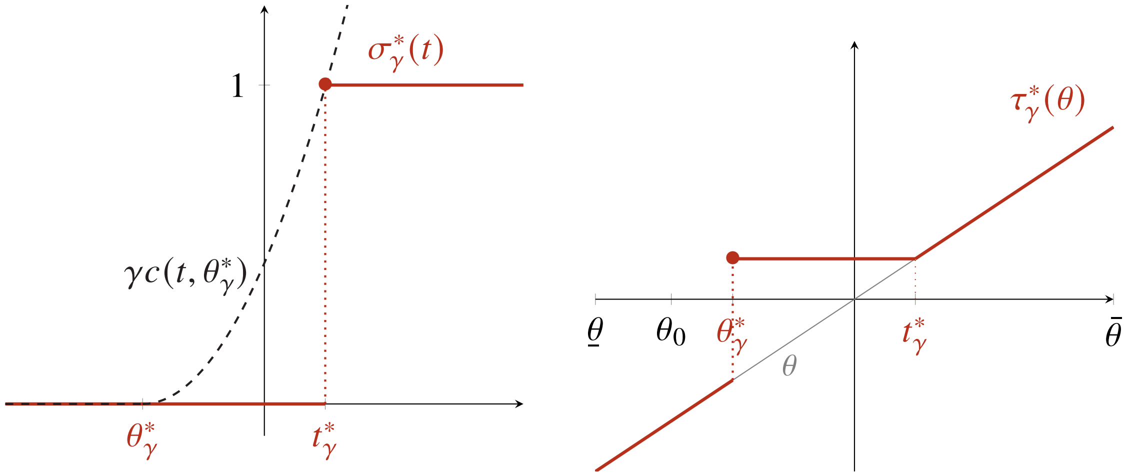
Optimal pass-fail rule. Optimizing the designer's payoff within the class of pass-fail mechanisms is much simpler than our original problem (P). Instead of solving an infinite dimensional program it reduces to the selection of a one dimensional allocation cutoff $t^\dagger$. In the next proposition we characterize the optimal mechanism within the class of pass-fail mechanisms.

Proposition 1. *Let t^*_γ be the unique solution to the equation $\psi(t) = 0$, where:*

$$\psi(t) = \begin{cases} t - \dfrac{F(t) - F(\theta(t))}{f(\theta(t))} & \text{if } t \in [0, \bar{\theta}[\\ \\ t - \dfrac{1 - F(\theta(t))}{f(\theta(t))} & \text{if } t \in [\bar{\theta}, \bar{\theta} + \sqrt{2/\gamma}] \end{cases} .$$

Then, the t^*_γ-pass-fail mechanism, denoted $(\sigma^*_\gamma, \tau^*_\gamma)$, is optimal in the class of pass-fail mechanisms. Moreover, under assumption 1 and 2, the optimal allocation cutoff t^*_γ belongs to the interval $]0, \sqrt{2/\gamma}[$, and the indifferent type $\theta^*_\gamma = \theta(t^*_\gamma)$ belongs to $]\theta_0, 0[$, for any $\gamma > 1/c(0, \underline{\theta})$.

Proof. See appendix A.1. □



(a) Optimal pass-fail selection rule.

(b) Implemented investment.

Figure 1.2: The optimal pass-fail allocation when $f(\theta) = \rho e^{-\rho\theta}/(e^{-\rho\underline{\theta}} - e^{-\rho\bar{\theta}})$ with $\rho = 2$, and $\gamma = 1$.

We illustrate the optimal pass-fail mechanism in figure 1.2 in the case of exponentially distributed types and a cost scaling factor normalized to one. Agents with an initial type given by θ^*_γ are indifferent between investing at the cutoff t^*_γ and staying at θ^*_γ. All agents below do not invest because they have too high a cost to reach t^*_γ and are excluded by the designer. Agents in the interval $[\theta^*_\gamma, t^*_\gamma]$ bunch at the allocation threshold, since it is the least costly final type that guarantees to be allocated the good with probability one. Agents with an initial type above t^*_γ, in turn, are already guaranteed to be allocated the good without any investment and thus keep their initial types at no cost.

The optimal allocation cutoff t^*_γ is strictly above the preference threshold of the designer. It is therefore optimal for the designer to commit to rejecting types above its preference threshold with probability one. This commitment on the part of the designer benefits him by encouraging a sufficiently large mass of agents to invest in a strictly positive type in equilibrium.

Optimal selection rule. We now state our main result. We show that the mechanism exhibited in proposition 1 is not only optimal in the class of pass-fail mechanisms but solves the designer's program (P).

Theorem 1. *If assumption 1 is satisfied, then* $(\sigma^*_\gamma, \tau^*_\gamma)$ *is a solution to* (P).

We defer the proof of theorem 1 to section 1.4. The main intuition for the result is the following. Any optimal mechanism must balance two conflicting forces acting on the designer's expected payoff. On the one hand, random monotone

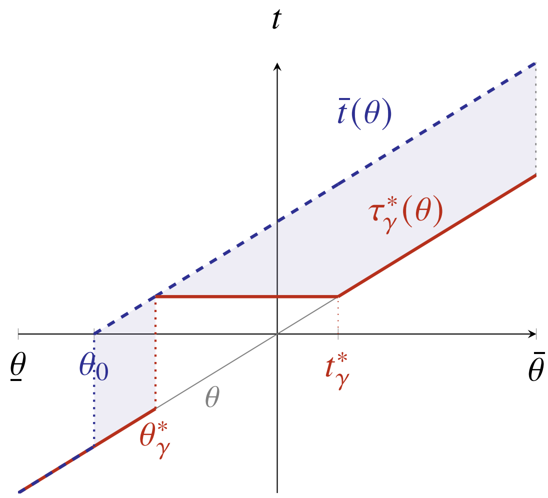


Figure 1.3: First-best (dashed lines) vs. second-best investment rule (plain lines)

selection rules incite agents' with initial types already above the selection cutoff to invest in higher final types, which benefits the designer. On the other hand, it also increases the probability to unduly reject some agents having invested in a final types above the designer's preference threshold, and decreases the mass of agents bunching at the selection cutoff. These two effects harm the designer's expected payoff. theorem 1 establishes that under assumption 1 the negative effect is always the strongest.

1.3.2. Properties of the optimal allocation

In this section, we describe the properties of the optimal selection rule. We first show that the optimal allocation is induces rationing compared to the first-best solution. Next, we perform comparative statics with respect to the cost scaling parameter γ.

Comparison to the first-best solution. We emphasized that it is optimal for the designer to commit ex-ante to rejecting types in the interval $[0, t^*_\gamma]$. In comparison, if the designer could observe the initial type of the agents, she could implement the first-best optimum by rejecting agents that do not invest in the maximum feasible final type given their cost, given by $\bar{t}(\theta) = \max\{t \in T \mid \gamma c(t, \theta) \leq 1\} = \theta + \sqrt{2/\gamma}$. The designer would thus admit all agents whose initial type lies in the interval $[\theta_0, \bar{\theta}]$ and implement their maximum possible level of investment. As a result, all the agents in the interval $[\theta_0, \theta^*_\gamma]$ are rationed under the second-best. Moreover, the designer has to forgo the value $\int_{\theta_0}^{\bar{\theta}} (\bar{t}(\theta) - \tau^*_\gamma(\theta)) f(\theta)\, d\theta$ due to the decrease in investments. This discrepancy is represented on figure 1.3.

Comparative statics. We first establish a preliminary result. We prove that if the distribution of types is sufficiently decreasing, then the optimal allocation cutoff never exceeds the highest initial type whatever the magnitude of investment costs.

Lemma 2. *If $f(\underline{\theta}) > 1/\bar{\theta}$, then the optimal cutoff t^*_γ belongs to the interval $[0, \bar{\theta}]$ for any $\gamma > 0$.*

Proof. See appendix A.3.1. □

The condition $f(\underline{\theta}) > 1/\bar{\theta}$ ensures that a sufficiently high mass of agents is concentrated at the bottom of the type distribution and thus guarantees that t^*_γ is bounded above by $\bar{\theta}$ for any $\gamma > 0$. The reason for this is that the mass effect always dominates the incentive effect for the designer when the type distribution is very decreasing. In other words, fixing a threshold of approval higher than $\bar{\theta}$ would always exclude a too large mass of low type agents compared to the gain in terms of type investment at the cutoff t^*_γ.

We conduct the comparative statics exercise under the assumption that $f(\underline{\theta}) \geq 1/\bar{\theta}$ in the main text. This assumption is made for readability and our results could easily be extended when it is not satisfied. The designer's optimal expected payoff is given by

$$V^*_\gamma = t^*_\gamma\big(F(t^*_\gamma) - F(\theta^*_\gamma)\big) + \int_{t^*_\gamma}^{\bar{\theta}} \theta f(\theta)\,\mathrm{d}\theta$$

Moreover, the agents' interim optimal payoff is given by:

$$U^*_\gamma(\theta) = \begin{cases} 0 & \text{if } \theta \in [\underline{\theta}, \theta^*_\gamma[\\ 1 - \gamma c(t^*_\gamma, \theta) & \text{if } \theta \in [\theta^*_\gamma, t^*_\gamma[\\ 1 & \text{if } \theta \in [t^*_\gamma, \bar{\theta}] \end{cases} .$$

First, we show that the designer's optimal allocation cutoff becomes looser as the investment costs increase.

Proposition 2. *The optimal cutoff t^*_γ defined in proposition 1 is monotonically decreasing towards zero as γ increases, while the minimal initial type being admitted θ^*_γ is monotonically increasing towards zero as γ increases. Moreover, the cutoffs t^*_γ and θ^*_γ are both converging to zero as γ tends to infinity.*

Proof. See appendix A.3.2. □

The insight for proposition 2 is as follows. Under the optimal mechanism, the agents with initial type θ^*_γ are indifferent between being rejected and keeping their initial type at cost $\gamma c(\theta^*_\gamma, \theta^*_\gamma) = 0$, or investing in a the final type type t^*_γ at a cost

$\gamma c(t^*_\gamma, \theta^*_\gamma) = 1$. Increasing the magnitude of the investment cost γ has the effect of increasing the marginal type θ^*_γ thereby excluding a larger mass of agents with low initial types. The optimal response of the designer to mitigate this reduction in the mass of approved agents at the bottom is to lower the admission threshold to restore higher investment incentives for those agents. proposition 2 has the following direct consequence: When the cost of investment becomes arbitrarily large for all agents, the optimal mechanism is for no agent to invest in a new type and for the designer to approve only positive initial types.

Corollary 1 (Asymptotics)**.** *The optimal experiment and optimal outcome respectively converges to* $\mathbb{1}\{t \geq 0\}$ *and to* $\mathbb{1}\{\theta \geq 0\}$ *as* $\gamma \to +\infty$*. Accordingly, the optimal payoffs converge to:*

$$V^*_\gamma \underset{\gamma \to +\infty}{\longrightarrow} \int_0^{\bar{\theta}} \theta f(\theta)\,\mathrm{d}\theta,$$

and:

$$U^*_\gamma \underset{\gamma \to +\infty}{\longrightarrow} 1 - F(0).$$

Corollary 1 confirms that when the investment cost becomes infinitely large for all agents, the optimal mechanism is the one that would be optimal if the designer could only use an approval rule, that is, if he solved the following problem:

$$\max_{\sigma\colon \Theta \to [0,1]} \int_{\underline{\theta}}^{\bar{\theta}} \theta\, \sigma(\theta) f(\theta)\,\mathrm{d}\theta.$$

Proposition 2 allows us to derive comparative statics on the designer's welfare as well as on the agents' welfare when γ increases.

Proposition 3 (Comparative statics)**.** *Under the optimal mechanism* $(\sigma^*_\gamma, \tau^*_\gamma)$ *the following claims are satisfied:*

(i) The designer's welfare V^*_γ *is decreasing in* γ*.*

(ii) There exists a cutoff $\tilde{\theta}_\gamma \in \,]\theta^*_\gamma, t^*_\gamma[$ *such that the interim welfare of agents* $U^*_\gamma(\theta)$ *is decreasing for all* $\theta \in [\theta^*_\gamma, \tilde{\theta}_\gamma]$ *and increasing for all* $\theta \in [\tilde{\theta}_\gamma, t^*_\gamma]$*.*

Quite naturally, proposition 3 establishes that the designer's payoff decreases as the investment becomes costlier. First, the allocation cutoff decreases as γ increases, which implies that agents who bunch at the threshold invest in a lower final type type in equilibrium. Moreover, the minimal type being approved θ^*_γ increases as γ grows. Hence, the total mass of agents being approved under the

optimal mechanism decreases as the cost of investment grows. Both effects lower the optimal payoff of the designer.

An upward scaling in the investment cost, however, has a non monotone effect on the agents' interim payoff. Indeed, the change in the agents' interim welfare due to a marginal increase in the investment cost can be decomposed in two effects which go in opposite directions. First, the direct effect of an increase in γ is to scale up the cost of investment for all types. Second, the indirect effect of an increase in γ is to lower the admission standard t^*_γ which decreases the distance from any θ to the allocation cutoff t^*_γ. This decreases the cost of investment for all agents. proposition 3 shows that there exists a type threshold $\tilde{\theta}_\gamma$ above which all agents are gaining from an increase in γ and below which all types are losing. The reason is that the indirect effect is stronger than the direct effect for all types which are sufficiently close to the allocation cutoff t^*_γ.

Comparative statics with respect to the welfare of agents turns out to be more intricate. Formally, the agents' aggregate welfare is given by:

$$U^*_\gamma = \underbrace{1 - F(\theta^*_\gamma)}_{\substack{\text{ex-ante allocation}\\ \text{probability}}} \underbrace{- \gamma \int_{\theta^*_\gamma}^{t^*_\gamma} c(t^*_\gamma, \theta) f(\theta)\, \mathrm{d}\theta}_{\text{aggregate investment cost}} .$$

The effect of γ on agents' welfare is ambiguous. The direct effect of an increase in γ is also to scale up the investment cost for all types of agents. Increasing γ also makes θ^*_γ increase, so the ex-ante allocation probability is decreasing. However, an increase in γ also decreases the allocation cutoff t^*_γ. This decreases the ex-ante probability of incurring the cost for agents at the top, and decreases the cost from investing to the cutoff for bunching agents. Depending on which of the effects dominate, the welfare of agents might increase or decrease. This suggests that the welfare of agents is not monotonic as a function of the magnitude of the investment costs.

1.4. Proof of theorem 1

In this section we state the main steps of the proof for theorem 1. Additional proofs can be found in appendix A.2. We first investigate the consequences of the monotonicity constraint imposed by lemma 1 for implementable investment rules. Next, we provide a characterization of incentive-compatible mechanisms in terms of a transformation of the agents' indirect utility function that we call the pseudo-utility function. Thanks to this characterization, we show that the problem

of the designer can be restated as a problem of calculus of variations where the optimization variable is the agents' pseudo-utility function. We prove that the objective functional of this variational program is an upper-semicontinuous convex functional on a compact and convex set. As a result, there must exist some extreme point of the domain that is a solution of the designer's problem. We provide necessary conditions on the shape of these extreme points, which, together with the tangent inequality for convex functionals, allows us to establish the optimality of the pass-fail selection rules.

Admissible mechanisms: definition. Lemma 1 sharpens the set of the investment rules that can be implemented by the designer. When the selection rule is monotone, no agent ever invests in a type that is strictly lower than its initial type, since this could only decrease its allocation probability. Therefore, we must have $\tau(\theta) \geq \theta$ for any $\theta \in \Theta$. We now show that there also exists an upper bound on agents' investments. Let us define $\bar{t}(\theta) = \max\{t \in T \,|\, \gamma c(t,\theta) \leq 1\}$. The type $\bar{t}(\theta)$ corresponds to the maximal type an agent with initial type θ would be willing to invest in, if he were allocated the good for sure. Since $c(\cdot,\theta)$ is a continuous and non-decreasing function over T for any $\theta \in \Theta$, the upper bound $\bar{t}(\theta)$ is defined implicitly by the following equation of t:

$$\gamma c(t,\theta) = 1, \tag{UB}$$

which has a closed-form solution given by $\bar{t}(\theta) = \theta + \sqrt{2/\gamma}$ for any $\theta \in \Theta$. Hence, we always have $\theta \leq \tau(\theta) \leq \bar{t}(\theta)$ for all $\theta \in \Theta$ and we can normalize T to $[\underline{\theta}, \bar{t}(\bar{\theta})]$ without loss of generality.

Any monotone selection rule must also induce agents approved with non-zero probability to acquire a non-negative final type. Let σ be a monotone selection rule and let $t^{\dagger}$ be the lowest final type type guaranteeing a strictly positive approval probability to the agents under σ, i.e.:

$$t^{\dagger} = \min\left\{t \in T \,|\, \sigma(t) > 0\right\}.$$

Since σ is monotone, $t^{\dagger}$ must be non-negative. Let $\theta^{\dagger}$ be the initial type defined as the solution to the following equation of θ:

$$\sigma(t^{\dagger}) = \gamma c(t^{\dagger}, \theta).$$

Denoting $\sigma(t^\dagger)$ by $\sigma^\dagger$, we obtain:

$$\theta^\dagger = t^\dagger - \sqrt{\frac{2\sigma^\dagger}{\gamma}}.$$

Clearly, $\theta^\dagger$ corresponds to the lowest type investing in a non-zero final type under the selection rule σ. In particular, an agent with type $\theta^\dagger$ is always indifferent between keeping type $\theta^\dagger$ and acquiring final type $t^\dagger$. Hence, we always have $\tau(\theta) = \theta$ for all $\theta \in [\theta_0, \theta^\dagger[$ and $\tau(\theta) \geq t^\dagger \geq 0$ for all $\theta \in [\theta^\dagger, \bar{\theta}]$. Remark that, differently from θ_0, the cutoff $\theta^\dagger$, together with the probability $\sigma^\dagger$, are part of the design and thus endogenous.

Third, let (σ, τ) be a mechanism such that σ is monotone and assume that there exists a type θ such that $\tau(\theta) = \bar{t}(\theta)$. Since τ satisfies (IC), it must be non-decreasing by Topkis' theorem. Therefore, there must exist a type $\tilde{\theta} \in \Theta$ such that $\tilde{\theta} = \min\{\theta \in \Theta \,|\, \tau(\theta) = \bar{t}(\theta)\}$. The condition (UB) then implies that $\sigma(\bar{t}(\tilde{\theta})) = 1$. Hence, since σ is increasing and bounded above by 1, it must also be that $\sigma(t) = 1$ for all $t \geq \bar{t}(\tilde{\theta})$. As a consequence, the final type $\bar{t}(\tilde{\theta})$ must be the smallest needed to pass with probability 1 under selection rule σ. Firstly, it might be the case that $\bar{t}(\tilde{\theta}) < \bar{\theta}$. Then, since τ satisfies (IC), all agents born with types in the interval $[\tilde{\theta}, \bar{t}(\tilde{\theta})[$ must invest exactly at the threshold $\bar{t}(\tilde{\theta})$ whereas all agents born with types in $[\bar{t}(\tilde{\theta}), \bar{\theta}]$ are already guaranteed to pass with probability 1 so $\tau(\theta) = \theta$ for all $\theta \geq \bar{t}(\tilde{\theta})$. Conversely, it might be the case that $\bar{t}(\tilde{\theta}) \geq \bar{\theta}$. This means that no initial type is initially guaranteed to pass with certain probability under σ and, again by (IC), all agents born with types in between $[\tilde{\theta}, \bar{\theta}]$ must invest at $\bar{t}(\tilde{\theta})$. We sum up the previous discussion in the following corollary.

Corollary 2. *If σ satisfies properties (i) and (ii) from lemma 1, then:*

(i) $\theta \leq \tau(\theta) \leq \bar{t}(\theta)$ *for all* $\theta \in [\theta_0, \bar{\theta}]$*;*

(ii) there always exists $\theta^\dagger \in [\theta_0, \bar{\theta}]$ *such that* $\tau(\theta) = \theta$ *for all* $\theta \in [\theta_0, \theta^\dagger[$ *and* $\tau(\theta) \geq 0$ *for all* $\theta \in [\theta^\dagger, \bar{\theta}]$*, and;*

(iii) if there exists $\tilde{\theta} \in [\theta_0, \bar{\theta}]$ *such that* $\tau(\tilde{\theta}) = \bar{t}(\tilde{\theta})$ *then we have the following:*

  (a) if $\bar{t}(\tilde{\theta}) < \bar{\theta}$*, then* $\tau(\theta) = \bar{t}(\tilde{\theta})$ *for all* $\theta \in [\tilde{\theta}, \bar{t}(\tilde{\theta})[$ *and* $\tau(\theta) = \theta$ *for all* $\theta \in [\bar{t}(\tilde{\theta}), \bar{\theta}]$*;*

  (b) if $\bar{t}(\tilde{\theta}) \geq \bar{\theta}$*, then* $\tau(\theta) = \bar{t}(\tilde{\theta})$ *for all* $\theta \in [\bar{t}(\tilde{\theta}), \bar{\theta}]$.

Accordingly, we say that any mechanism (σ, τ) satisfying (IC) as well as all properties from lemma 1 and corollary 2 is *admissible*.

Admissible mechanisms: characterization. The indirect utility of an agent of type θ under some selection rule σ is given by

$$U(\theta) = \max_{t \in T} \ \sigma(t) - \gamma c(t, \theta).$$

Given the quadratic form of the cost function, simple algebra on the agents' objective function reveals that

$$U(\theta) = \gamma \left(\max_{t \in T} \left\{ t\theta - \left(\frac{t^2}{2} - \frac{\sigma(t)}{\gamma} \right) \right\} - \frac{\theta^2}{2} \right).$$

Accordingly, we say that the value function defined by

$$u(\theta) = \max_{t \in T} \ t\theta - \left(\frac{t^2}{2} - \frac{\sigma(t)}{\gamma} \right),$$

is the agent's *pseudo-utility* function under the rule σ. Let $\underline{u}$ be the pseudo-utility function induced by the selection rule $\underline{\sigma}(t) = 0$ and $\bar{u}$ be the pseudo-utility function induced by the selection rule $\bar{\sigma}(t) = \mathbb{1}\{t \geq 0\}$.[12] We have:

$$\underline{u}(\theta) = \frac{\theta^2}{2}$$

and

$$\bar{u}(\theta) = \begin{cases} \frac{1}{\gamma} & \text{if } \theta \in [\theta_0, 0[\\ \frac{1}{\gamma} + \frac{\theta^2}{2} & \text{if } \theta \in [0, \bar{\theta}] \end{cases}.$$

The next proposition characterizes admissible mechanisms in terms of properties of the pairs (u, τ).

Lemma 3. *A mechanism (σ, τ) is admissible if, and only if, the pseudo-utility function u and investment τ induced by σ satisfy the following properties:*

(i) u is a convex function over $[\theta_0, \bar{\theta}]$ and is hence differentiable almost everywhere in that interval. Moreover, the envelope formula $u'(\theta) = \tau(\theta)$ must hold at all points where u is differentiable;

(ii) $\underline{u}(\theta) \leq u(\theta) \leq \bar{u}(\theta)$, for all $\theta \in [\theta_0, \bar{\theta}]$;

(iii) $\theta \leq u'(\theta) \leq \bar{t}(\theta)$ for almost all $\theta \in [\theta_0, \bar{\theta}]$;

[12] Remark that $\underline{\sigma}$ and $\bar{\sigma}$ are respectively the lowest and greatest recommendation rules in the class of monotone selection rules.

(iv) $u(\theta) + u'(\theta)^2/2 - \theta u'(\theta) \leq 1/\gamma$ *for almost all* $\theta \in [\theta_0, \bar{\theta}]$*;*

(v) There always exists $\theta^\dagger \in [\theta_0, \bar{\theta}]$ *such that* $u(\theta) = \underline{u}(\theta)$ *for all* $\theta \in [\theta_0, \theta^\dagger]$ *and* u *is increasing and strictly above* $\underline{u}(\theta)$ *for all* $\theta \in]\theta^\dagger, \bar{\theta}]$*;*

(vi) If there exists $\hat{\theta} \in [\theta_0, \bar{\theta}]$ *such that* $u'(\hat{\theta}) = \bar{t}(\hat{\theta})$*, then, if* $\bar{t}(\hat{\theta}) < \bar{\theta}$ *we must have:*

$$u(\theta) = \begin{cases} \underline{u}(\theta) & \text{if } \theta \in [\theta_0, \hat{\theta}[\\ \underline{u}(\hat{\theta}) + \bar{t}(\hat{\theta})(\theta - \hat{\theta}) & \text{if } \theta \in [\hat{\theta}, \bar{t}(\hat{\theta})] \\ \bar{u}(\theta) & \text{if } \theta \in]\bar{t}(\hat{\theta}), \bar{\theta}] \end{cases},$$

and if $\bar{t}(\hat{\theta}) \geq \bar{\theta}$ *then:*

$$u(\theta) = \begin{cases} \underline{u}(\theta) & \text{if } \theta \in [\theta_0, \hat{\theta}[\\ u(\theta) = \underline{u}(\hat{\theta}) + \bar{t}(\hat{\theta})(\theta - \hat{\theta}) & \text{if } \theta \in [\hat{\theta}, \bar{\theta}] \end{cases}$$

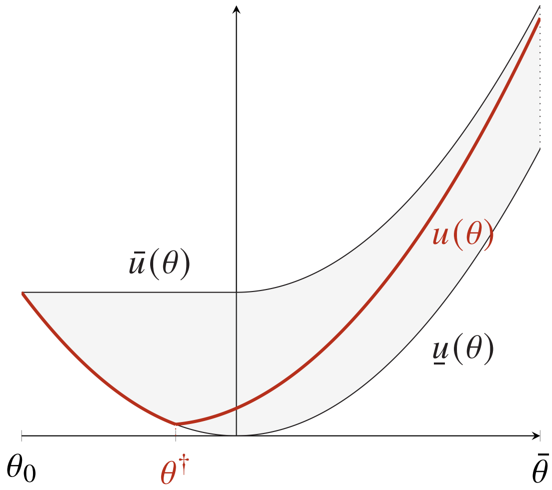
Proof. See appendix A.2.2 □

Property (i) is a consequence of Proposition 2 in Rochet (1987) and ensures that the pair (u, τ) is consistent with the first-order conditions of the agents' optimization program wherever σ is differentiable.[13] By the envelope formula, the function u' is non-decreasing, hence the function u must be convex. Another way to see why u must be convex is the following fact. If we let $\varphi(t) = t^2/2 - \sigma(t)/\gamma$, then:

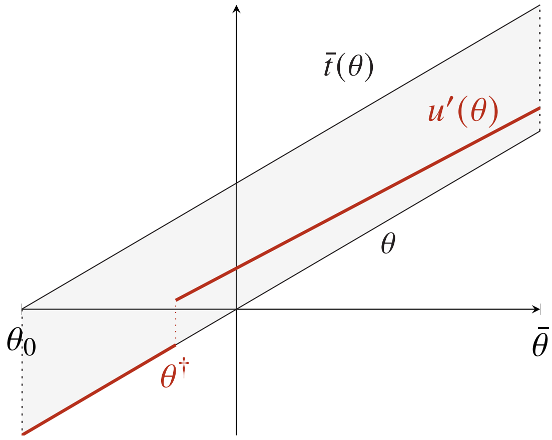
$$u(\theta) = \max_{t \in T} \; \theta t - \varphi(t).$$

The function u thus corresponds to the Legendre-Fenchel transform of the function φ, which is convex by definition. The previous expression also highlights the fact that u corresponds to the indirect utility an agent with quasilinear payoff would get in a monopolistic screening problem (Mussa and Rosen, 1978) under some tariff φ. Note, however, that unlike standard screening problems, the function φ must be bounded between $\underline{\varphi}(t) = t^2/2 - \underline{\sigma}(t)/\gamma$ and $\bar{\varphi}(t) = t^2/2 - \bar{\sigma}(t)/\gamma$ because of the probability constraint on σ and of lemma 1. This implies property (ii). Property (iii), in turn, is simply the translation in terms of the pseudo-utility function of corollary 2. Let us now explain property (iv). For this, we remind first that the *outcome* $x\colon [\theta_0, \bar{\theta}] \to [0, 1]$ of the selection rule σ is given by $x(\theta) = \sigma(\tau(\theta))$ for all $\theta \in [\theta_0, \bar{\theta}]$. Using property (i), we can rewrite the outcome as a function of u and u' by substituting $\tau(\theta)$ by $u'(\theta)$ almost everywhere. Indeed, remembering

[13]We know that, as a non-decreasing function, σ must be differentiable on T except maybe on a countable subset of points.



(a) An admissible pseudo-utility u.



(b) Its induced investment rule u'.

Figure 1.4: An admissible pseudo-utility function $u \in \mathcal{U}$. Its induced investment rule is given by its derivative u'. Here $u(\theta) = 0.45\times(\theta+0.4)^2+0.1\times(\theta+0.4)+0.08$.

that $U(\theta) = \gamma(u(\theta) - \theta^2/2)$ and remarking that $\sigma(\tau(\theta)) = U(\theta) + \gamma c(\tau(\theta), \theta)$ we obtain:

$$x(\theta) = \gamma\left(u(\theta) + \frac{u'(\theta)^2}{2} - \theta u'(\theta)\right),$$

for almost every $\theta \in [\theta_0, \bar{\theta}]$. Thus, property (iv) expresses the fact that x must be bounded above by 1. Importantly, it is easy to show that for any u satisfying properties (ii) and (iii), we have $x(\theta) \geq 0$ almost everywhere. Nevertheless, we need to add property (iv) in the characterization as there exists some functions satisfying (ii) and (iii) such that $x(\theta) > 1$ for some θ in $[\theta_0, \bar{\theta}]$. Finally, properties (v) and (vi) are direct consequences of corollary 2.

Variational program. Using the previous characterization, we can first define the following set.

Definition 4. *The* feasible set *of pseudo-utility functions is defined by:*

$$\mathcal{U} = \left\{u\colon [\theta_0, \bar{\theta}] \to \mathbb{R} \,\middle|\, u \text{ satisfies properties (ii) to (vi) from lemma 3}\right\}.$$

Any $u \in \mathcal{U}$ can be generated by some admissible mechanism (σ, τ) and any admissible mechanism (σ, τ) induces some pseudo-utility $u \in \mathcal{U}$. We thus refer to any $u \in \mathcal{U}$ as an admissible pseudo-utility. For the purpose of illustration, an instance of admissible pseudo-utility $u \in \mathcal{U}$ is depicted on figure 1.4a while its induced investment rule, given by its derivative, is depicted on figure 1.4b. Second, we can also rewrite the integrand of the designer's objective as a function of the

agents' pseudo-utility by substituting $\tau(\theta)$ by $u'(\theta)$ and $\sigma(\tau(\theta))$ by $x(\theta)$ almost everywhere:

$$\tau(\theta)\,\sigma(\tau(\theta)) = \gamma u'(\theta)\left(u(\theta) + \frac{u'(\theta)^2}{2} - \theta u'(\theta)\right),$$

for almost every $\theta \in [\theta_0, \bar{\theta}]$. Therefore, for any pseudo-utility function $u \in \mathcal{U}$, the designer's payoff is given by:

$$V(u) = \int_{\theta_0}^{\bar{\theta}} \Lambda\big(\theta, u(\theta), u'(\theta)\big)\,\mathrm{d}\theta.$$

where, for any $\theta \in [\theta_0, \bar{\theta}]$, the function $\Lambda(\theta, \cdot, \cdot)$ is defined by

$$\Lambda(\theta, x, y) = \gamma y\left(x + \frac{y^2}{2} - \theta y\right) f(\theta)$$

for any $(x, y) \in [\underline{u}(\theta), \bar{u}(\theta)] \times [\theta, \bar{t}(\theta)]$. The problem of the designer can therefore be expressed as the following problem of *calculus of variations*[14]:

$$\max_{u \in \mathcal{U}} \int_{\theta_0}^{\bar{\theta}} \Lambda\big(\theta, u(\theta), u'(\theta)\big)\,\mathrm{d}\theta \tag{V}$$

Unfortunately, we cannot apply the standard resolution methods for program (V) because the objective functional does not satisfy the necessary conditions to use the first-order approach (see, for instance Clarke, 2013, Chapter 14, Section 1). Therefore, we have to resort to different techniques to solve it.

Parameterization. First of all, we parametrize the problem (V) with respect to $\theta^\dagger$, the type whereupon the function u starts taking-off with a non-negative slope from the lower bound $\underline{u}$. Consider the following set of functions:

$$\begin{aligned}\mathcal{U}(\theta^\dagger) = \Big\{u\colon [\theta^\dagger, \bar{\theta}] \to \mathbb{R} \,\Big|\, & u \text{ is convex and increasing,}\\ & \underline{u}(\theta) \le u(\theta) \le \bar{u}(\theta),\\ & \theta \le u'(\theta) \le \bar{t}(\theta),\\ & u(\theta) + u'(\theta)^2/2 - \theta u'(\theta) \le 1/\gamma,\\ & u(\theta^\dagger) = \underline{u}(\theta^\dagger)\Big\}\end{aligned}$$

[14]We refer to Clarke (2013) for a thorough treatment of calculus of variations.

for any $\theta^\dagger \in [\theta_0, \bar{\theta}]$. Remark that for any $u \in \mathcal{U}(\theta^\dagger)$ we have $\Lambda(\theta, \underline{u}(\theta), \underline{u}'(\theta)) = 0$ for all $\theta \in [\theta_0, \theta^\dagger]$ and $\Lambda(\theta, u(\theta), u'(\theta)) \geq 0$ for all $\theta \in]\theta^\dagger, \bar{\theta}]$. We can therefore always integrate the objective starting from $\theta^\dagger$. Accordingly, we define the parameterized objective functional $V_{\theta^\dagger}$ as follows:

$$V_{\theta^\dagger}(u) = \int_{\theta^\dagger}^{\bar{\theta}} \Lambda\big(\theta, u(\theta), u'(\theta)\big) \mathrm{d}\theta.$$

Restriction to extreme points of $\mathcal{U}(\theta^\dagger)$. We first prove the two following crucial observations.

Lemma 4. *For any $\theta^\dagger \in [\theta_0, \bar{\theta}]$, the set $\mathcal{U}(\theta^\dagger)$ is convex and is compact with respect to the supremum-norm.*

Proof. See appendix A.2.3. □

Lemma 5. *For any $\theta^\dagger \in [\theta^\dagger, \bar{\theta}]$, the functional $V_{\theta^\dagger} : \mathcal{U}(\theta^\dagger) \longrightarrow \mathbb{R}$ is* upper semicontinuous. *Moreover, if assumption 1 is satisfied, then it is also* convex.

Proof. See appendix A.2.4. □

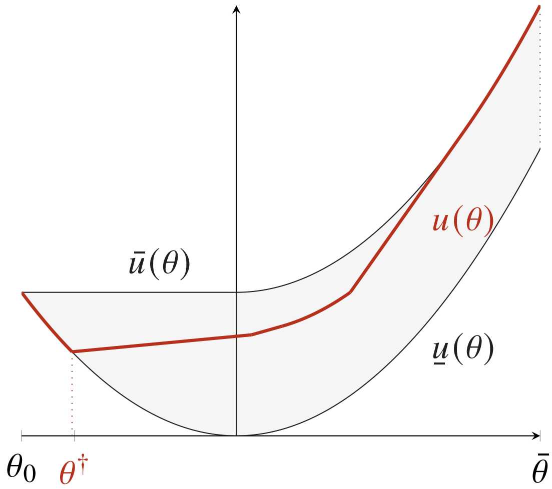
Lemmas 4 and 5 have the following implications: First, by the *Krein–Milman theorem* (Aliprantis and Border, 2006, Theorem 7.68) it must be that the set $\mathcal{U}(\theta^\dagger)$ is the closed, convex hull of its extreme points and, in particular, that the set $\mathcal{E}(\theta^\dagger)$ of extreme points of $\mathcal{U}(\theta^\dagger)$ is non-empty. Second, *Bauer's Maximum Principle* (Aliprantis and Border, 2006, Theorem 7.69) ensures that the functional $V_{\theta^\dagger}$ must admit a maximizer which belongs to $\mathcal{E}(\theta^\dagger)$.

A function $u \in \mathcal{U}(\theta^\dagger)$ is an extreme point if there does not exist $u_1, u_2 \in \mathcal{U}$ and $\alpha \in]0, 1[$ such that $u = \alpha u_1 + (1 - \alpha) u_2$. We provide next an equivalent and more convenient definition of extreme points.

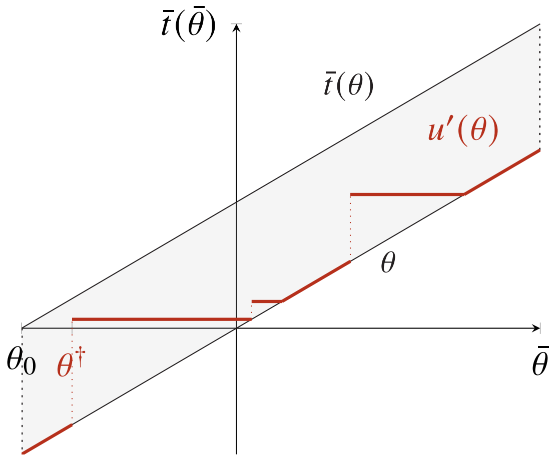
Definition 5. *Let C be a compact and convex subset of a locally convex topological vector space X. A point $x \in C$ is an extreme point if, and only if, for every direction $h \in X$ such that $h \neq 0$, we either have that $x + h \notin C$ or $x - h \notin C$, or both.*

We proved in lemma 4 that $\mathcal{U}(\theta^\dagger)$ is a compact and convex subset of the normed linear space $(\mathcal{C}([\theta^\dagger, \bar{\theta}]), \|\cdot\|_\infty)$. Therefore, a function u belongs to $\mathcal{E}(\theta^\dagger)$ if, and only if, there does not exist any direction $h \in \mathcal{C}([\theta^\dagger, \bar{\theta}])$ such that $u - h \in \mathcal{U}$ and $u + h \in \mathcal{U}$. In the next lemma, we provide necessary conditions on the extreme points of $\mathcal{U}(\theta^\dagger)$.

Lemma 6. *If $u \in \mathcal{U}(\theta^\dagger)$ is an extreme point then:*



(a) An extreme point u.



(b) Its induced investment rule u'.

Figure 1.5: An extreme point $u \in \mathcal{E}(\theta^\dagger)$ together with its induced investment rule u'.

(i) u must increasing and piecewise affine on the interval $[\theta^\dagger, 0]$, i.e., there exists a countable set I and a sequence $(a_i, b_i)_{i\in I}$ such that $0 \leq a_i \leq a_{i+1}$ and $b_i \in \mathbb{R}$ for all $i \in I$, and that $u(\theta) = \max\{a_i\theta + b_i \,|\, i \in I, a_0\theta^\dagger + b_0 = \underline{u}(\theta^\dagger)\}$ for all $[\theta^\dagger, 0[$;

(ii) u must be a convex connection between increasing affine and quadratic arcs on the interval $]0, \bar{\theta}]$, i.e., there exists a countable set J, a sequence $(c_j)_{j\in J}$ such that $u(0) \leq c_j \leq c_{j+1} \leq 1/\gamma$ for any $j \in J$, as well as a collection of intervals $([\underline{x}_j, \bar{x}_j[)_{j\in J}$ such that $[\underline{x}_j, \bar{x}_j[\subseteq]0, \bar{\theta}]$, that $\bar{x}_j \leq \underline{x}_{j+1}$ for all $j \in J$, and that $u(\theta) = \theta^2/2 + c_j$ for every $j \in J$ and any $\theta \in [\underline{x}_j, \bar{x}_j[$. Moreover, whenever $\theta \in]0, \bar{\theta}] \setminus \cup_{j\in J}[\underline{x}_j, \bar{x}_j[$, u must be increasing and piecewise affine, and must join two quadratic arcs while staying convex.

Proof. See appendix A.2.5 □

Lemma 6 provides necessary conditions on the extreme points. It shows that if $u \in \mathcal{E}(\theta^\dagger)$, then it must be the convex junction between increasing affine and quadratic arcs. An example of extreme point is depicted on figure 1.5a together with its derivative on figure 1.5b. The reason why extreme points take this particular form is that such curves either make the bounds on the derivative binding when u is confounded with θ or $\bar{t}(\theta)$, or make the convexity constraint binding when u' is flat, i.e., u is affine. However, due to the constraint that $u'(\theta) \geq 0$, the function u cannot be quadratic if $\theta < 0$ because otherwise one would have $u'(\theta) = \theta < 0$.

Lemma 6 has an important implication. When searching for an optimal solution of (V), one can restrict attention without loss of optimality to functions that are

convex connections of linear and quadratic arcs. Let us define the function $u^\dagger$ given by:

$$u^\dagger(\theta) = \begin{cases} \underline{u}(\theta^\dagger) + \bar{t}(\theta^\dagger)(\theta - \theta^\dagger) & \text{if } \theta \in [\theta^\dagger, \bar{t}(\theta^\dagger)[\\ \bar{u}(\theta) & \text{if } \theta \in [\bar{t}(\theta^\dagger), \bar{\theta}] \end{cases} .$$

This function corresponds to the pseudo-utility function implemented by a $\bar{t}(\theta^\dagger)$-pass-fail rule restricted to the interval $[\theta^\dagger, \bar{\theta}]$. Indeed, remember that lemma 3 implies that $(u^\dagger)'(\theta)$ corresponds to the optimal investment after the cutoff $\theta^\dagger$. It is easy to verify that:

$$(u^\dagger)'(\theta) = \begin{cases} \bar{t}(\theta^\dagger) & \text{if } \theta \in [\theta^\dagger, \bar{t}(\theta^\dagger)[\\ \theta & \text{if } \theta \in [\bar{t}(\theta^\dagger), \bar{\theta}] \end{cases} \tag{1.3}$$

if $\theta^\dagger < \bar{\theta} - \sqrt{2/\gamma}$, and

$$(u^\dagger)'(\theta) = \bar{t}(\theta^\dagger), \tag{1.4}$$

if $\theta^\dagger \geq \bar{\theta} - \sqrt{2/\gamma}$. This corresponds to the investment rule that would be implemented under a $\bar{t}(\theta^\dagger)$-pass-fail rule. Importantly, the function $u^\dagger$ satisfies all the necessary conditions in lemma 6 and is an extreme point of $\mathcal{U}(\theta^\dagger)$. Also importantly, $u^\dagger$ bounds any other function $u \in \mathcal{U}(\theta^\dagger)$ from above so $u^\dagger - u \geq 0$ for any $u \in \mathcal{U}(\theta^\dagger)$. The final step of our proof is to show that $u^\dagger$ is an optimal solution.

To do so, we first recall the following characterization of convex functionals.

Lemma 7 (Above the tangent property for convex functions)**.** *Let X be a normed space, C be a non-empty closed convex subset of X and $\varphi\colon C \longrightarrow \mathbb{R}$ be a Gâteaux differentiable function, with Gâteaux derivative at $x \in C$ in direction $h \in X$ given by* $\mathrm{D}\varphi(x)(h)$*. Then, φ is convex if, and only if:*

$$\varphi(y) \geq \varphi(x) + \mathrm{D}\varphi(x)(y - x)$$

for all $(x, y) \in C \times C$.

By lemma 4, we know that the set $\mathcal{U}(\theta^\dagger)$ is a compact and convex subset of the normed linear space $(\mathcal{C}([\theta^\dagger, \bar{\theta}]), \|\cdot\|_\infty)$. Moreover, we also proved in lemma 5 that $V_{\theta^\dagger}$ is convex over $\mathcal{U}(\theta^\dagger)$ under assumption 1. We now prove that the functional $V_{\theta^\dagger}$ is Gâteaux differentiable everywhere on $\mathcal{U}(\theta^\dagger)$ and we give a closed form for its Gâteaux derivative.

Lemma 8. *The functional $V_{\theta^\dagger}$ is Gâteaux differentiable and has a Gâteaux*

derivative at u in direction h given by:

$$\mathrm{D}V_{\theta^\dagger}(u)(h) = \int_{\theta^\dagger}^{\bar{\theta}} \Bigg(-\Bigg(\Big(u(\theta) + \frac{1}{2}u'(\theta)^2 - \theta u'(\theta) \Big) + u'(\theta)\,(u'(\theta) - \theta) \Bigg) f'(\theta)$$
$$+ \Big(u'(\theta) - 2u''(\theta)\,(u'(\theta) - \theta) - u'(\theta)\,(u''(\theta) - 1) \Big) f(\theta) \Bigg) h(\theta)\,\mathrm{d}\theta.$$

Proof. See appendix A.2.6. □

Hence, lemma 7 and lemma 8 together imply that:

$$V_{\theta^\dagger}(v) \geq V_{\theta^\dagger}(u) + \mathrm{D}V_{\theta^\dagger}(u)(v - u) \tag{1.5}$$

for any $(u, v) \in \mathcal{U}(\theta^\dagger) \times \mathcal{U}(\theta^\dagger)$ and any $\theta^\dagger \in [\theta_0, \bar{\theta}]$. In particular, equation (1.5) must be satisfied when $v = u^\dagger$ and when u is an extreme point, since $\mathcal{E}(\theta^\dagger) \subset \mathcal{U}(\theta^\dagger)$. That is:

$$V_{\theta^\dagger}(u^\dagger) \geq V_{\theta^\dagger}(u) + \mathrm{D}V_{\theta^\dagger}(u)(u^\dagger - u) \tag{1.6}$$

for any $u \in \mathcal{E}(\theta^\dagger)$. We now prove the following important result.

Lemma 9. *If assumption 1 is satisfied, then* $\mathrm{D}V_{\theta^\dagger}(u)(u^\dagger - u) \geq 0$ *for any* $u \in \mathcal{E}(\theta^\dagger)$.

Proof. See appendix A.2.7. □

Lemma 9 is the final step of our proof. Indeed, equation (1.6) and lemma 9 together entail that $V_{\theta^\dagger}(u^\dagger) \geq V_{\theta^\dagger}(u)$ for any $u \in \mathcal{E}(\theta^\dagger)$, proving the optimality of $u^\dagger$. The proof of lemma 9 relies on the fact that at any extreme point $u \in \mathcal{E}(\theta^\dagger)$, we have either $u''(\theta) = 0$ and $u'(\theta) = a$ for some constant $a > 0$ on intervals where u is linear, or $u''(\theta) = 1$ and $u'(\theta) = \theta$ on intervals where u is quadratic. This, together with the fact that $u^\dagger(\theta) - u(\theta) \geq 0$ for any $\theta \in [\theta_0, \bar{\theta}]$ implies that the integrand in the expression of the Gâteaux derivative given in lemma 8 is always positive. Intuitively, this condition means that at a given cutoff $\theta^\dagger$, when we restrict to the extreme points of the domain $\mathcal{U}(\theta^\dagger)$, deviating to the pseudo-utility $u^\dagger$ always locally increase the value of the principal's objective. Note that any extreme point $u \in \mathcal{E}(\theta^\dagger)$ can be implemented by an increasing step selection rule. The linear parts of u then correspond to regions where agents bunch at the next step and the quadratic parts correspond to regions where agents have an investment cost too large to reach the next step and therefore keep their type at zero cost. lemma 9 therefore basically states that among all step selection rules, the best one is the one with two steps. A step with value zero and a step with value one, separated by the

allocation cutoff $\bar{t}(\theta^\dagger)$. Optimizing the designer's payoff over the one-dimensional parameter $\theta^\dagger$ ends the proof of theorem 1.

1.5. Extensions

We study three extensions of our model. First, we add a capacity constraint to the principal allocation problem. Second, we solve the problem of a utilitarian social planner. In both cases, the optimal selection rule remains pass-fail. Finally, we relax the principal's commitment power. Instead of committing to a mechanism, the principal bases her allocation decision on the information provided by an intermediary with aligned preferences. We show that the principal-optimal allocation can be implemented by the intermediary through information design.

1.5.1. Capacity constrained designer

In the baseline model, the designer possesses the same mass of resources than the total mass of agents. In this extension, we assume instead that the designer is *capacity constrained*. That is, she can at most allocate a positive measure $\kappa < 1$, of resources. Under that additional constraint, the problem of the designer writes as follows:

$$\begin{aligned} \underset{\sigma,\tau}{\text{maximize}} \quad & \int_{\underline{\theta}}^{\bar{\theta}} \tau(\theta)\, \sigma(\tau(\theta)) f(\theta)\, \mathrm{d}\theta \\ \text{subject to} \quad & \tau(\theta) \in \arg\max_{t \in T} \sigma(t) - \gamma c(t,\theta) \qquad \text{(IC)} \\ \text{and} \quad & \int_{\underline{\theta}}^{\bar{\theta}} \sigma(\tau(\theta)) f(\theta)\; \mathrm{d}\theta \leq \kappa \qquad \text{(C)} \end{aligned}$$

Letting $\lambda \geq 0$ be the Lagrange multiplier on the constraint (C), we can observe that the previous problem reduces to the same problem than (P) where the preference threshold of the designer has been moved from 0 to λ:

$$\begin{aligned} \underset{\sigma,\tau,\lambda}{\text{maximize}} \quad & \int_{\underline{\theta}}^{\bar{\theta}} (\tau(\theta) - \lambda)\, \sigma(\tau(\theta)) f(\theta)\, \mathrm{d}\theta + \lambda\kappa \\ \text{subject to} \quad & \tau(\theta) \in \arg\max_{t \in T} \sigma(t) - \gamma c(t,\theta) \qquad \text{(IC)} \end{aligned}$$

Since the structure of the problem is unchanged, we can use the proof of theorem 1 verbatim on the proviso that the designer restrict herself to selection rules that are zero below the preference threshold λ, and non-decreasing above. We thus have

the following result.

Proposition 4. *Pass-fail selection rules solve the capacity constrained allocation problem. Moreover, if the capacity constraint is binding, then the optimal threshold is tighter than under no capacity constraint.*

This proposition follows from the standard Kuhn and Tucker method, and we therefore omit its proof. The idea is as follows. If the solution to the constrained program is the same as in the case when $\kappa = 1$ then the multiplier λ is null and the capacity constraint is not binding at the optimum. If the solution differs, then $\lambda > 0$ and the capacity constraint kicks in. Whenever it is the case, the selection cutoff must, by definition, weakly increase compared to the unconstrained selection cutoff. Otherwise, it would mean that the designer allocates at least the same mass of resources than under the unconstrained program, a contradiction.

1.5.2. Utilitarian welfare

We now consider the problem of a utilitarian social planner seeking to maximize weighted social welfare. Formally, the planner's program writes as follows:

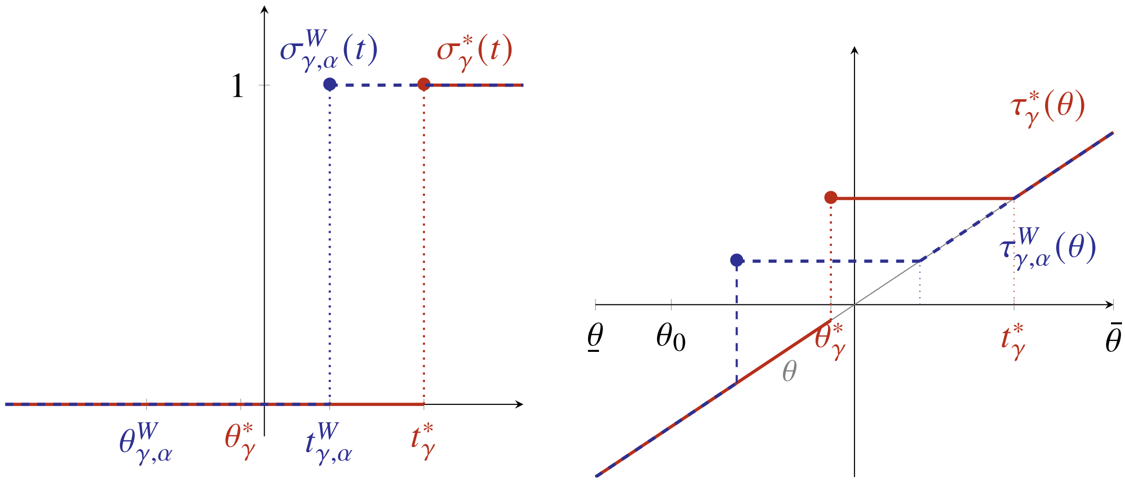
$$\begin{aligned} \underset{\sigma,\tau}{\text{maximize}} \quad & \int_{\underline{\theta}}^{\bar{\theta}} \tau(\theta)\, \sigma(\tau(\theta)) f(\theta)\, \mathrm{d}\theta + \alpha \int_{\underline{\theta}}^{\bar{\theta}} \big(\sigma(\tau(\theta)) - \gamma c(\tau(\theta), \theta)\big) f(\theta)\, \mathrm{d}\theta \\ \text{subject to} \quad & \tau(\theta) \in \arg\max_{t \in T} \sigma(t) - \gamma c(t, \theta) \qquad \text{(IC)} \end{aligned}$$

where $\alpha > 0$ is the Pareto weight the planner assigns to the welfare of agents. When $\alpha < 1$, the planner cares more about the welfare of the principal, and conversely when $\alpha > 1$. We prove that the welfare-optimal selection rules are also deterministic, and have a smaller allocation threshold compared to the principal-optimal rule.

Proposition 5. *The welfare-optimal pass-fail rule solves the problem of the planner. Moreover, the welfare-optimal allocation cutoff is lower than under the optimal selection rule of the principal.*

Proof. See appendix A.4. □

Proposition 5 states that taking into account the aggregate welfare of agents does not affect the deterministic structure of the optimal selection rule, and leads to less severe selection than under the principal-optimal selection rule. We illustrate the discrepancy between the welfare-optimal and the principal-optimal selection rules in figure 1.6. Naturally, taking into account the investment costs of agents



(a) Optimal pass-fail selection rule. (b) Implemented investment.

Figure 1.6: The welfare-optimal pass-fail allocation (dashed lines) vs. the principal-optimal allocation (plain lines), when $f(\theta) = \rho e^{-\rho\theta}/(e^{-\rho\underline{\theta}} - e^{-\rho\bar{\theta}})$ with $\rho = 0.2$, $\gamma = 1$, and $\alpha = 1$.

pushes the welfare-optimal allocation cutoff to the left of the principal-optimal one.

Unlike proposition 4, the proof of proposition 5 requires some adaptations compared to the proof of theorem 1. The reason is that the planner cannot restrict itself without loss to monotonic mechanisms as described in lemma 1. Indeed, the planner, unlike the principal, may have an interest in accepting negative types if the Pareto weight α is high enough. Nevertheless, the constraint (IC) still guarantees that the selection rule σ is a non-decreasing function over the domain T. The proof relies only on adapting the characterization of implementable pseudo-utility functions under such allocation functions, but otherwise works in the same way as the proof of theorem 1. Indeed, when the planner's program is rewritten in variational form, the agents' welfare term is linear in the pseudo-utility, thus not affecting the convexity of the objective functional.

1.5.3. Implementation through information design

In this section, we consider a slightly modified setup. There are three players: the receiver, the sender and the agents. The receiver decides whether to allocate the resource to the agents, $a \in A = \{0, 1\}$. The agents choose whether to invest in a new type. Their payoffs are the same than in the baseline model. We let $\pi \in \Delta(\Theta)$ denote the prior probability measure on agents' types. An investment strategy for an agent is a stochastic mapping $\tau\colon \Theta \to \Delta(T)$ associating any initial types θ to a conditional distribution $\tau(\theta)$ over final types. For any state $\theta \in \Theta$, we define the

interim expected cost of agent θ as:

$$C(\tau, \theta) = \gamma \int_T c(t, \theta)\, \tau(\mathrm{d}t \mid \theta).$$

The sender provides information to the receiver by committing to a *statistical experiment* (σ, S) (Blackwell, 1951, 1953), which consists in an endogenously chosen set of signal realizations S and a stochastic mapping $\sigma\colon T \to \Delta(S)$ associating any realized final type type t to a probability distribution $\sigma(t)$ over S. We denote by σ the collection $(\sigma(t))_{t\in T}$. We also assume that the sender has the same payoff than the receiver.

Given an experiment (σ, S), when he anticipates the investment strategy of any agent of type θ to be $\tau(\theta)$, designer's posterior belief on an agents' type whose signal realization is s, is given by be Bayes rule:

$$\mu_{\sigma,\tau}(\tilde{T} \mid s) = \frac{\displaystyle\int_{\tilde{T}\times\Theta} \sigma(s \mid t)\, \tau(\mathrm{d}t \mid \theta)\, \pi(\mathrm{d}\theta)}{\displaystyle\int_{T\times\Theta} \sigma(s \mid t)\, \tau(\mathrm{d}t \mid \theta)\, \pi(\mathrm{d}\theta)},$$

for any Borel set $\tilde{T} \subseteq T$ and any $s \in \bigcup_{t\in T} \mathrm{supp}(\sigma(t))$. Let

$$\hat{t}_{\sigma,\tau}(s) = \int_T t\, \mu_{\sigma,\tau}(\mathrm{d}t \mid s).$$

be the associated the receiver's posterior expectation over one agents' final type conditional on signal s.

An allocation strategy for the receiver is a stochastic mapping $\alpha\colon S \to \Delta(A)$ associating any signal realization s to a probability distribution $\alpha(s)$ over allocation decisions. With a slight abuse of notation, we let $\alpha\colon S \to [0, 1]$ be the measurable function such that $\alpha(s) = \alpha(1 \mid s)$ for any signal realization $s \in S$. For any anticipated strategy of the agent τ, the receiver's optimal strategy α must maximize the expected posterior mean type conditional on allocation, given by

$$\int_{S\times T\times\Theta} \hat{t}_{\sigma,\tau}(s)\, \alpha(s)\, \sigma(\mathrm{d}s \mid t)\, \tau(\mathrm{d}t \mid \theta)\, \pi(\mathrm{d}\theta).$$

Let us define the set:

$$S(\sigma, \tau) = \left\{ s \in S \;\middle|\; \hat{t}_{\sigma,\tau}(s) \geq 0 \right\},$$

for any (σ, τ). The receiver's optimal strategy under any σ and τ is thus given by

$$\alpha_{\sigma,\tau}(s) = \mathbb{1}_{S(\sigma,\tau)}(s),$$

for all $s \in S$.

Given an experiment (σ, S), the interim probability that the receiver allocates the object under anticipation τ, and agent's strategy is τ', is

$$\rho_{\sigma,\tau}(\tau', \theta) = \int_{S\times T} \alpha_{\sigma,\tau}(s)\, \sigma(\mathrm{d}s \mid t)\, \tau'(\mathrm{d}t \mid \theta),$$

for every $\theta \in \Theta$. Thus, the agents' interim payoff when the designer anticipates τ but the actual investment strategy is τ' is given by:

$$U_{\sigma,\tau}(\tau', \theta) = \rho_{\sigma,\tau}(\tau', \theta) - C(\tau', \theta).$$

We thus say that τ is *agent-incentive-compatible* if it is a best response to the receiver's optimal strategy under experiment (σ, S). That is:

$$U_{\sigma,\tau}(\tau, \theta) \geq U_{\sigma,\tau}(\tau', \theta), \tag{A-IC}$$

for any $\theta \in \Theta$ and any τ'.

Let us define the sender's equilibrium payoff under experiment (σ, S) as follows:

$$V(\sigma, \tau) = \int_{S\times T\times\Theta} \hat{t}_{\sigma,\tau}(s)\, \alpha_{\sigma,\tau}(s)\, \sigma(\mathrm{d}s \mid t)\, \tau(\mathrm{d}t \mid \theta)\, \pi(\mathrm{d}\theta),$$

The problem for designer is to find an experiment (σ, S) that maximizes her ex-ante expected payoff given that τ must be agent-incentive-compatible under (σ, S), that is:

$$\underset{\sigma,\tau}{\text{maximize}}\ V(\sigma, \tau) \text{ subject to (A-IC).}$$

Akin to Kamenica and Gentzkow (2011), we prove a recommendation principle entailing that the choice of an optimal experiment (σ, S) can be reduced to the choice of an allocation recommendation rule $\sigma\colon T \to \Delta(A)$. The proof follows similar steps as in Perez-Richet and Skreta (2022b) and goes as follows: Start from an arbitrary experiment (σ, S) under which the equilibrium is $(\alpha_{\sigma,\tau}, \tau)$. Then, define the garbled experiment ς which pools together all signals leading designer to allocate under the original experiment σ. The experiment induces the designer to follow the recommendations and also maintains the same interim allocation probabilities for the agents so no type has any incentive to deviate from

its investment strategy under ς. Hence, garbling the original experiment leads to the same payoffs for the designer as well as agents. Although the result that following the recommendations under the new experiment is standard, the result that τ remains a best response to ς is distinctive of our framework. We state the recommendation principle more formally in the next lemma together with its proof.

Lemma 10 (Recommendation principle)**.** *Fix an experiment* (σ, S) *and let* τ *be* (A-IC)*. Consider the experiment* (ς, A) *defined by*

$$\varsigma(1 \mid t) = \sigma\big(S(\sigma, \tau) \,\big|\, t\big)$$

for every $t \in T$*. Then, all the following properties are satisfied:*

(i) The receiver always follows the sender's recommendations in equilibrium, i.e., $\alpha_{\varsigma,\tau}(1) = 1$*;*

(ii) Interim allocation probabilities are the same under σ *and* ς*, i.e.,* $\rho_{\sigma,\tau}(\tau', \theta) = \rho_{\varsigma,\tau}(\tau', \theta)$ *for any* τ' *and* θ*;*

(iii) τ *is* (A-IC) *under* ς*;*

(iv) Equilibrium payoffs are the same under σ *and* ς*, i.e.,* $U_{\sigma,\tau}(\tau, \theta) = U_{\varsigma,\tau}(\tau, \theta)$ *for any* θ *and* $V(\sigma, \tau) = V(\varsigma, \tau)$*.*

Proof. See appendix A.5. □

With a slight abuse of notation, we let $\sigma \colon T \to [0, 1]$ be the measurable function such that $\sigma(t) = \sigma(1 \mid t)$ for all $t \in T$. Henceforth, we refer to the function σ as sender's recommendation rule. Under that formulation, the ex-ante expected payoff for the sender is given by:

$$V(\sigma, \tau) = \int_{T\times\Theta} t\sigma(t)\, \tau(\mathrm{d}t \mid \theta)\, \pi(\mathrm{d}\theta).$$

In turn, the interim payoff for the agent is given by:

$$U_\sigma(\tau, \theta) = \int_T \sigma(t)\, \tau(\mathrm{d}t \mid \theta) - C(\tau, \theta).$$

When choosing a recommendation rule, the designer must make sure that it is individually rational for the receiver to follow its recommendation but also that it is agent-incentive-compatible. The sender's recommendation rule is receiver-individually-rational if the receiver's expected payoff conditional on an allocation

recommendation is non-negative, i.e., $\int_{T\times\Theta} t\sigma(t)\,\tau(\mathrm{d}t \mid \theta)\,\pi(\mathrm{d}\theta) \geq 0$, and her expected payoff following a non-allocation recommendation is non-positive, i.e., $\int_{T\times\Theta} t(1-\sigma(t))\,\tau(\mathrm{d}t \mid \theta)\,\pi(\mathrm{d}\theta) \leq 0$. Combining these two inequalities, we obtain:

$$V(\sigma,\tau) \geq \max\left\{0, \int_{T\times\Theta} t\,\tau(\mathrm{d}t \mid \theta)\,\pi(\mathrm{d}\theta)\right\}, \tag{R-IR}$$

The sender's problem thus consists in finding the recommendation rule solving:

$$\underset{\sigma,\tau}{\text{maximize}}\ V(\sigma,\tau) \text{ subject to (R-IR) and (A-IC).}$$

We now argue that the constraint (R-IR) in sender's problem can be removed without loss of generality. Indeed, a non-informative experiment achieves the lower bound required by (R-IR) and satisfies the (A-IC) constraint. A non-informative experiment induces a constant recommendation rule, thus $\int_{T\times\Theta} t\,\tau(\mathrm{d}t \mid \theta)\,\pi(\mathrm{d}\theta) = \hat{\theta}_\pi$ since there is no way for agents to affect the allocation probability by investing. If $\hat{\theta}_\pi \geq 0$ then sender fixes $\sigma(t) = 1$ for every t and if $\hat{\theta}_\pi < 0$ then she fixes $\sigma(t) = 0$ for every t. Hence, designer's payoff is $\max\{0, \hat{\theta}_\pi\}$ which corresponds to the lower bound in (R-IR). But, remark that any recommendation rule solving the relaxed program

$$\underset{\sigma,\tau}{\text{maximize}}\ V(\sigma,\tau) \text{ subject to (A-IC)} \tag{P}$$

must give the designer at least the value than under an uninformative experiment and thus satisfy constraint (R-IR). Slightly abusing of notation, we henceforth denote by $\tau(\theta)$ the designer's preferred selection of the correspondence $\mathcal{T}(\theta) = \arg\max_{t\in T} \sigma(t) - \gamma c(t,\theta)$. Under that formulation, the optimization program of the designer can be stated as our original allocation problem:

$$\begin{aligned} \underset{\sigma,\tau}{\text{maximize}} \quad & \int_{\underline{\theta}}^{\bar{\theta}} \tau(\theta)\,\sigma(\tau(\theta)) f(\theta)\,\mathrm{d}\theta \\ \text{subject to} \quad & \tau(\theta) \in \arg\max_{t\in T}\ \sigma(t) - \gamma c(t,\theta) \end{aligned} \tag{A-IC}$$

Interestingly, this equivalence implies that the designer has no additional value when she has more commitment power.

Proposition 6. *Commitment to a mechanism has no additional value than commitment to an experiment only to the designer.*

1.6. Conclusion

We study the optimal design of non-market allocation mechanisms that take into account agents' productive investment incentives. In our baseline model, a principal has a unit mass of resources to allocate to a unit mass of agents. The agents are characterized by a type, which is the only payoff-relevant variable for the principal. Agents undertake costly investments, the outcome of which is a type transformation observable by the principal. The principal wishes to allocate the good to agents whose types are above some ideal threshold. She commits ex-ante to an selection rule, contingent on the outcome of the type improvement. Our main result states that pass-fail selection rules are optimal, under the assumption that the preference threshold of the principal lies in the right tail of the distribution of agents' initial types and that the agents' investment costs are increasing and convex in the amount of their investment.

We also cover three extensions of the model. First, we consider a capacity constrained designer. Then we consider the problem faced by a utilitarian social planner. In both cases, we show that pass-fail rules remain optimal. The optimal cutoff increases when the capacity constraint is binding. Conversely, taking into account agents' costs when maximizing social welfare leads the planner to choose a lower threshold than in the baseline solution. Finally, we show that the optimal allocation can be implemented by an information designer. This implies that weakening the principal's commitment power does not reduce her optimal payoff.

There are several interesting directions for future work. The first, which is the most natural but nevertheless challenging, is to extend the characterization of optimal mechanisms to more general distributions and cost functions. Second, characterizing selection rules that would combine productive investment incentives together with the possibility that agents might engage in falsification or, equivalently, in costly signaling, is also a promising avenue. Finally, adding affirmative action constraints to our problem, e.g., in the form of quotas, is also an important extension, whose resolution would allow for the design of fairer resource allocation mechanisms.

Bibliography

Aghion, P., M. Dewatripont, and P. Rey (1994): "Renegotiation Design with Unverifiable Information," *Econometrica*, 62, 257–282.

Akbarpour, M., P. Dworczak, and S. D. Kominers (2020): "Redistributive Allocation Mechanisms," *SSRN Electronic Journal*.

Akbarpour, M., S. D. Kominers, K. M. Li, S. Li, and P. Milgrom (2022): "Investment Incentives in Truthful Approximation Mechanisms," *Working paper*.

Aliprantis, C. D. and K. C. Border (2006): *Infinite Dimensional Analysis*, Berlin Heidelberg: Springer-Verlag, third ed.

Amador, M. and K. Bagwell (2020): "Money burning in the theory of delegation," *Games and Economic Behavior*, 121, 382–412.

Ambrus, A. and G. Egorov (2017): "Delegation and nonmonetary incentives," *Journal of Economic Theory*, 171, 101–135.

Arozamena, L. and E. Cantillon (2004): "Investment Incentives in Procurement Auctions," *Review of Economic Studies*, 71, 1–18.

Ashlagi, I., F. Monachou, and A. Nikzad (2021a): "Costly Signaling with Heterogeneous Outside Options," *Working paper*.

——— (2021b): "Optimal Dynamic Allocation: Simplicity through Information Design," *SSRN Electronic Journal*.

Athey, S. and I. Segal (2013): "An Efficient Dynamic Mechanism," *Econometrica*, 81, 2463–2485.

Ball, I. (2019): "Scoring Strategic Agents," *arXiv*.

Ben-Porath, E., E. Dekel, and B. L. Lipman (2014): "Optimal Allocation with Costly Verification," *American Economic Review*, 104, 3779–3813.

Bergemann, D. and P. Strack (2022): "Progressive participation," *Theoretical Economics*, 17, 1007–1039.

Bergemann, D. and J. Välimäki (2002): "Information Acquisition and Efficient Mechanism Design," *Econometrica*, 70, 1007–1033.

Bhaskar, D. and E. D. Sadler (2020): "Resource Allocation with Positive Externalities," *SSRN Electronic Journal*.

Blackwell, D. (1951): "Comparison of Experiments," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, ed. by J. Neyman, Berkeley: University of California Press, 93–102.

——— (1953): "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, 24, 265–272.

Börgers, T., D. Krähmer, and R. Strausz (2015): *An Introduction to the Theory of Mechanism Design*, New York: Oxford University Press.

Carlier, G. (2001): "A general existence result for the principal-agent problem with adverse selection," *Journal of Mathematical Economics*, 35, 129–150.

Chakravarty, S. and T. R. Kaplan (2013): "Optimal allocation without transfer payments," *Games and Economic Behavior*, 77, 1–20.

Chan, J. and E. Eyster (2003): "Does Banning Affirmative Action Lower College Student Quality?" *American Economic Review*, 93, 858–872.

Chung, T.-Y. (1991): "Incomplete Contracts, Specific Investments, and Risk Sharing," *Review of Economic Studies*, 58, 1031–1042.

Clarke, E. H. (1971): "Multipart pricing of public goods," *Public Choice*, 11, 17–33.

Clarke, F. (2013): *Functional Analysis, Calculus of Variations and Optimal Control*, vol. 264 of *Graduate Texts in Mathematics*, London: Springer-Verlag.

Condorelli, D. (2012): "What money can't buy: Efficient mechanism design with costly signals," *Games and Economic Behavior*, 75, 613–624.

——— (2013): "Market and non-market mechanisms for the optimal allocation of scarce resources," *Games and Economic Behavior*, 82, 582–591.

Condorelli, D. and B. Szentes (2020): "Information Design in the Holdup Problem," *Journal of Political Economy*, 128, 681–709.

Daley, B., M. Schwarz, and K. Sonin (2012): "Efficient investment in a dynamic auction environment," *Games and Economic Behavior*, 75, 104–119.

Daskalakis, C., A. Deckelbaum, and C. Tzamos (2017): "Strong Duality for a Multiple-Good Monopolist," *Econometrica*, 85, 735–767.

Dilmé, F. (2019): "Pre-trade private investments," *Games and Economic Behavior*, 117, 98–119.

Erlanson, A. and A. Kleiner (2019): "A note on optimal allocation with costly verification," *Journal of Mathematical Economics*, 84, 56–62.

Fang, H. and A. Moro (2011): "Theories of Statistical Discrimination and Affirmative Action: A Survey," in *Handbook of Social Economics*, Elsevier B.V., vol. 1, 133–200.

Frankel, A. and N. Kartik (2019): "Muddled Information," *Journal of Political Economy*, 127, 1739–1776.

——— (2022): "Improving Information from Manipulable Data," *Journal of the European Economic Association*, 20, 79–115.

Fryer, R. G. and G. C. Loury (2013): "Valuing Diversity," *Journal of Political Economy*, 121, 747–774.

Fryer, R. G., G. C. Loury, and T. Yuret (2007): "An Economic Analysis of Color-Blind Affirmative Action," *Journal of Law, Economics, and Organization*, 24, 319–355.

Gershkov, A., B. Moldovanu, P. Strack, and M. Zhang (2021): "A Theory of Auctions with Endogenous Valuations," *Journal of Political Economy*, 129, 1011–1051.

Grossman, S. J. and O. D. Hart (1986): "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94, 691–719.

Groves, T. (1973): "Incentives in Teams," *Econometrica*, 41, 617–631.

Gul, F. (2001): "Unobservable Investment and the Hold-Up Problem," *Econometrica*, 69, 343–376.

Halac, M. and P. Yared (2020): "Commitment versus Flexibility with Costly Verification," *Journal of Political Economy*, 128, 4523–4573.

Hart, O. and J. Moore (1988): "Incomplete Contracts and Renegotiation," *Econometrica*, 56, 755–785.

HARTLINE, J. D. AND T. ROUGHGARDEN (2008): "Optimal mechanism design and money burning," in *Proceedings of the fortieth annual ACM symposium on Theory of computing*, New York, NY, USA: ACM, 75–84.

HATFIELD, J. W., F. KOJIMA, AND S. D. KOMINERS (2014): "Investment Incentives in Labor Market Matching," *American Economic Review*, 104, 436–441.

——— (2019): "Strategy-Proofness, Investment Efficiency, and Marginal Returns: An Equivalence," *SSRN Electronic Journal*.

HATFIELD, J. W., F. KOJIMA, AND Y. NARITA (2016): "Improving schools through school choice: A market design approach," *Journal of Economic Theory*, 166, 186–211.

HU, L., N. IMMORLICA, AND J. W. VAUGHAN (2018): "The Disparate Effects of Strategic Manipulation," *arXiv*, 259–268.

KAMENICA, E. AND M. GENTZKOW (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

KATTWINKEL, D. (2020): "Allocation with Correlated Information: Too Good to be True," in *Proceedings of the 21st ACM Conference on Economics and Computation*, New York, NY, USA: ACM, vol. 224, 109–110.

KATTWINKEL, D. AND J. KNOEPFLE (2022): "Costless Information and Costly Verification: A Case for Transparency," *SSRN Electronic Journal*.

KATTWINKEL, D., A. NIEMEYER, J. PREUSSER, AND A. WINTER (2022): "Mechanisms without transfers for fully biased agents," *SSRN Electronic Journal*, 1–25.

KLEIN, B., R. G. CRAWFORD, AND A. A. ALCHIAN (1978): "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law and Economics*, 21, 297–326.

KLEINBERG, J. AND M. RAGHAVAN (2020): "Algorithmic classification and strategic effort," *ACM SIGecom Exchanges*, 18, 45–52.

KLEINER, A. AND A. MANELLI (2019): "Strong Duality in Monopoly Pricing," *Econometrica*, 87, 1391–1396.

KLEINER, A., B. MOLDOVANU, AND P. STRACK (2021): "Extreme Points and Majorization: Economic Applications," *Econometrica*, 89, 1557–1593.

Kováč, E. and T. Mylovanov (2009): "Stochastic mechanisms in settings without monetary transfers: The regular case," *Journal of Economic Theory*, 144, 1373–1395.

Laffont, J.-J. and J. Tirole (1986): "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 94, 614–641.

Lau, S. (2008): "Information and bargaining in the hold-up problem," *RAND Journal of Economics*, 39, 266–282.

MacLeod, W. B. and J. M. Malcomson (1993): "Investments, Holdup, and the Form of Market Contracts," *American Economic Review*, 83, 811–837.

Manelli, A. M. and D. R. Vincent (2007): "Multidimensional mechanism design: Revenue maximization and the multiple-good monopoly," *Journal of Economic Theory*, 137, 153–185.

Mussa, M. and S. Rosen (1978): "Monopoly and product quality," *Journal of Economic Theory*, 18, 301–317.

Myerson, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6, 58–73.

Mylovanov, T. and A. Zapechelnyuk (2017): "Optimal Allocation with Ex Post Verification and Limited Penalties," *American Economic Review*, 107, 2666–2694.

Niemeyer, A. and J. Preusser (2022): "Simple Allocation with Correlated Type," *Working paper*.

Onuchic, P. (2022): "Recent Contributions to Theories of Discrimination," *arXiv*.

Ortoleva, P., E. Safonov, and L. Yariv (2021): "Who Cares More? Allocation with Diverse Preference Intensities," Tech. rep., National Bureau of Economic Research, Cambridge, MA.

Ottaviani, M. (2021): "Grantmaking," *Working paper*.

Perez-Richet, E. and V. Skreta (2022a): "Fraud-proof non-market allocation mechanisms," *Working paper*.

——— (2022b): "Test Design Under Falsification," *Econometrica*, 90, 1109–1142.

Ray, D. and R. Sethi (2010): "A Remark on Color-Blind Affirmative Action," *Journal of Public Economic Theory*, 12, 399–406.

Riley, J. and R. Zeckhauser (1983): "Optimal Selling Strategies: When to Haggle, When to Hold Firm," *Quarterly Journal of Economics*, 98, 267–289.

Rochet, J.-C. (1987): "A necessary and sufficient condition for rationalizability in a quasi-linear context," *Journal of Mathematical Economics*, 16, 191–200.

Rochet, J.-C. and P. Choné (1998): "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, 66, 783–826.

Rogerson, W. P. (1992): "Contractual Solutions to the Hold-Up Problem," *Review of Economic Studies*, 59, 777.

Tirole, J. (1986): "Procurement and Renegotiation," *Journal of Political Economy*, 94, 235–259.

Tomoeda, K. (2019): "Efficient investments in the implementation problem," *Journal of Economic Theory*, 182, 247–278.

Vickrey, W. (1961): "Counterspeculation, auctions and competitive sealed tenders," *Journal of Finance*, 16, 8–37.

Weitzman, M. L. (1977): "Is the Price System or Rationing More Effective in Getting a Commodity to Those Who Need it Most?" *The Bell Journal of Economics*, 8, 517.

Williamson, O. E. (1979): "Transaction-Cost Economics: The Governance of Contractual Relations," *Journal of Law and Economics*, 22, 233–261.

Yoon, K. (2011): "Optimal mechanism design when both allocative inefficiency and expenditure inefficiency matter," *Journal of Mathematical Economics*, 47, 670–676.

2. Persuading a wishful thinker[1]

Abstract

We analyze a model of persuasion in which Receiver forms wishful non-Bayesian beliefs. The effectiveness of persuasion depends on Receiver's material stakes: it is more effective when intended to encourage risky behavior that potentially leads to a high payoff and less effective when intended to encourage more cautious behavior. We illustrate this insight with applications showing why informational interventions are often ineffective in inducing greater investment in preventive health treatments, how financial advisors might take advantage of their clients overoptimistic beliefs and why strategic information disclosure to voters with different partisan preferences can lead to belief polarization in an electorate.

2.1. Introduction

It is generally assumed in models of strategic communication that receivers update beliefs in a perfectly rational manner, as would a Bayesian statistician. Yet, a substantial literature in psychology and behavioral economics shows that the process by which individuals interpret information and form beliefs is not guided solely by a desire for accuracy but often depends on their motivations and material incentives. This phenomenon is generally referred to as *motivated inference* (Kunda, 1987, 1990), and a common manifestation of it is *wishful thinking*: the tendency of

[1]This chapter is a joint work with Daniel Barreto and previously circulated under the title "Wishful Thinking: Persuasion and Polarization." This version of the paper is the one submitted to the journal *Games & Economic Behavior* on February 28, 2022, currently at the "Reject and Resubmit" stage. We would like to point out that the version presented in this thesis is being extensively modified in view of the resubmission of the paper. We thank Jeanne Hagenbach and Eduardo Perez-Richet for their support. We also thank S. Nageeb Ali, Roland Bénabou, Michele Fioretti, Alexis Ghersengorin, Simon Gleyze, Emeric Henry, Deniz Kattwinkel, Frédéric Koessler, Laurent Mathevet, Meg Meyer, Daniel Monte, Nikhil Vellodi, Adrien Vigier and Yves Le Yaouanq for their valuable feedbacks and comments, as well as seminar audiences at Sciences Po, Paris School of Economics, São Paulo School of Economics (FGV) and at the Econometric Society European Meeting 2021. All remaining errors are ours. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 850996 – MOREV and 101001694 – IMEDMC).

individuals to let their *preferences about outcomes* influence the way they process information, leading to beliefs that are systematically biased towards outcomes they wish to be true.[2] In this paper we investigate how wishful thinking affects the effectiveness of persuasion, i.e., the probability or frequency with which a sender is able to induce a receiver to take her preferred action.

Following Caplin and Leahy (2019), we propose a model in which the receiver's belief updating rule is non-bayesian: after observing an informative signal, Receiver forms beliefs by trading off their anticipatory value against the psychological cost of distorting beliefs away from Bayesian ones. As a result, Receiver's beliefs are stakes-dependent, i.e., they depend on his preferences, and overweight the state associated with the highest payoff, giving rise to overoptimism.

Distortions in beliefs lead to distortions in Receiver's behavior: some actions end up being favored, meaning that they are taken more often (i.e., after the reception of a strictly greater set of possible signals) relative to a Bayesian decision-maker. When he only has two available actions, wishful thinking leads Receiver to favor the action associated with the highest payoff and the highest payoff variability. If one of the two actions induces the highest possible payoff and the other induces the highest payoff variability, then which of the two is favored depends on the magnitude of Receiver's belief distortion cost. As such, the effectiveness of information provision as a tool to incentivize agents might vary with individuals' material stakes: *persuasion is more effective when it is aimed at encouraging behavior that is risky but can potentially yield very high returns and less effective when it is aimed at encouraging more cautious behavior*. We illustrate this insight in applications in which wishful beliefs can play an important role.

Application 1: Information Provision and Preventive Health Care. In this application a public health agency designs an information policy about the risk of infection of an illness in order to promote a preventive treatment that can be adopted by individuals at some cost. Since not adopting the treatment is the action that can potentially yield the highest payoff (in case the illness is not severe) and also the action with the highest payoff variability, it is favored by wishful receivers. As such, information campaigns aimed at promoting preventive behavior are less effective. We also show how the effectiveness of information campaigns are impacted by the severity of the disease and the effectiveness of the treatment.

[2]There exists abundant experimental evidence of wishful thinking. See in particular Bénabou and Tirole (2016), page 150 and Benjamin (2019) Section 9, as well as, e.g., Weinstein (1980), Mijović-Prelec and Prelec (2010), Mayraz (2011), Heger and Papageorge (2018), Coutts (2019), Engelmann, Lebreton, Schwardmann, van der Weele, and Chang (2019) or Jiao (2020).

This application sheds light on the stylized fact that individuals are consistently investing too little in preventive health care treatments, even if offered at low prices (especially in developing countries, see Dupas, 2011; Chandra, Handel, and Schwartzstein, 2019; Kremer, Rao, and Schilbach, 2019, Section 3.1) and that informational interventions are often ineffective in inducing more investment in preventive health care devices (see, in particular, Dupas, 2011, Section 4, and Kremer et al., 2019, Section 3.3). Recent literature conjectures that individuals might not be responsive to such information campaigns because they prefer to hold optimistic prospects about their health risks (see Schwardmann, 2019 and Kremer et al., 2019, Section 3.3).[3] Our model formalizes this argument.

Application 2: Persuading a Wishful Investor. In this application, we consider the interaction between a financial broker and her potential client. The broker designs reports about the (continuously distributed) return of some risky financial product to persuade the client to buy the asset. We show that a financial broker interested in selling a risky product is always more effective when persuading a wishful investor.

This application formalizes why some professional financial advisors might sometimes not act in the best interest of their clients by making investment recommendations that take advantage of their biases and mistaken beliefs (see, for instance, Mullainathan, Noeth, and Schoar, 2012 or Beshears, Choi, Laibson, and Madrian, 2018, Section 9) as well as why some consulting firms seem to specialize in advice misconduct and cater to biased consumers (Egan, Matvos, and Seru, 2019). It also helps explaining why the online betting industry puts so much effort into persuasion. Indeed, Babad and Katz (1991) document that individuals generally display wishful thinking when they take part in lotteries: they prefer to think they will win and are therefore more receptive to information encouraging risky bets.

Application 3: Public Persuasion and Political Polarization. Belief polarization along partisan lines is a pervasive and much debated feature of contemporary societies. Although such polarization can be partly caused by differential access to information, evidence suggests that it is exacerbated by the fact that individuals tend to make motivated inferences about the *same* piece of information (Babad, 1995; Thaler, 2020).

[3]There exists compelling experimental evidence that such self-deception exists in the medical testing context (Lerman, Hughes, Lemon, Main, Snyder, Durham, Narod, and Lynch, 1998; Oster, Shoulson, and Dorsey, 2013; Ganguly and Tasoff, 2017).

In this application we explore the relationship between optimal information disclosure to wishful citizens and belief polarization. Following Alonso and Câmara (2016), we model a majority voting setting in which an electorate, differentiated in terms of partisan preferences, uses information disclosed by a politician to vote on a proposal. Wishful thinking leads voters with different preferences to adopt different beliefs after being exposed to a public signal: those voting against or for the proposal distort their beliefs in opposite directions, giving rise to polarization. Sender's optimal public experiment consists in persuading the median voter, which maximizes the number of voters distorting beliefs in opposite directions. We show that if partisan preferences are symmetrically distributed around the median, then Sender's optimal information policy generates maximal belief polarization in the electorate as a byproduct. This adds nuance to the argument that motivated thinking is one of the drivers of polarization: not only can motivated thinking lead to polarization, but the strategic disclosure of information to a motivated electorate can also accentuate this tendency[4].

Related literature. The persuasion and information design literature[5] has initially focused on the problem of influencing rational Bayesian decision-makers as in the seminal contributions of Kamenica and Gentzkow (2011) and Bergemann and Morris (2016). By introducing non-Bayesian updating in the form of motivated beliefs formation, we contribute to the literature studying persuasion of receivers subject to mistakes in probabilistic inferences.[6][7] Levy, Moreno de Barreda, and Razin (2018) analyze a Bayesian persuasion problem where a sender can send multiple signals to a receiver subject to correlation neglect. Benjamin, Bodoh-Creed, and Rabin (2019) provide an example of persuasion game where

[4]This application is related to the paper by Le Yaouanq (2021) who constructs a model of large elections with motivated voters. As in our model, the formation of motivated beliefs by citizens leads voters with different preferences to hold different beliefs after observing the same information. We find, as he does, that greater heterogeneity in partisan preferences increases belief polarization but has no effect on the policy implemented in equilibrium. This is, however, the consequence of a different modelling assumption. Namely, that information is endogenously designed to persuade the median voter, whose vote is not distorted relative to a Bayesian voter.

[5]See Bergemann and Morris (2019) and Kamenica (2019) for reviews of this literature.

[6]See Benjamin (2019) for a review of the literature. In particular, wishful thinking belongs to preference-biased inferences reviewed in Benjamin (2019), Section 9.

[7]It is interesting to note that an active literature also explores how errors in strategic reasoning (Eyster, 2019) affect equilibrium outcomes in strategic communication games. Although in our model Receiver understands all the strategic issues, we believe, nevertheless, that it is important to mention that players' misunderstanding of their strategic environment might also lead them to make errors in statistical inference even if they update beliefs via Bayes' rule, as in Mullainathan, Schwartzstein, and Shleifer (2008), Ettinger and Jehiel (2010), Hagenbach and Koessler (2020) and Eliaz, Spiegler, and Thysen (2021b,a) who consider communication games where players make inferential errors because of a coarse understanding of their environment.

Receiver exhibits base-rate neglect when updating beliefs. In de Clippel and Zhang (2020) the receiver holds subjective beliefs which belong to a broader class of distorted Bayesian posteriors. In contrast, in our model, Receiver's belief formation process optimally trades-off the benefits and costs associated with maintaining non-Bayesian beliefs as in the work of Caplin and Leahy (2019).

On the one hand, we assume that Receiver's value from maintaining inaccurate beliefs comes from the anticipation of the payoff he will achieve in equilibrium. Intuitively, it represents the idea that individuals might derive utility from the anticipation of future outcomes, be them good or bad. This hypothesis has been widely used in the literature to study how anticipatory emotions affect physical choices (see, e.g., Loewenstein, 1987; Caplin and Leahy, 2001) as well as choices of beliefs (Bénabou and Tirole, 2002; Brunnermeier and Parker, 2005; Bracha and Brown, 2012; Caplin and Leahy, 2019). Receiver's choice of beliefs is thus a way of satisfying his psychological need to be optimistic about the best-case outcomes or, on the contrary, to avoid the dread and anxiety associated with the worst-case outcomes. This hypothesis is supported experimentally by Engelmann et al. (2019), who find significant evidence that wishful thinking is caused by the desire to reduce anxiety associated with anticipating bad events. It is important to note that while anticipatory utility may be a strong motive for manipulating one's beliefs, it is not the only possible one. This differentiates wishful thinking from the more general concept of motivated reasoning, which is usually defined as the degree to which individuals' cognition is affected by their motivations.[8] Different motivations from anticipated payoffs have been explored in the literature such as cognitive dissonance avoidance (Akerlof and Dickens, 1982; Golman, Loewenstein, Moene, and Zarri, 2016), preference to believe in a "Just World" (Bénabou and Tirole, 2006), maintaining high motivation when individuals are aware of being subject to a form of time-inconsistency (Bénabou and Tirole, 2002, 2004) or satisfying the need to belong to a particular identity (Bénabou and Tirole, 2011).

On the other hand, we assume distorting beliefs away from the Bayesian benchmark is subject to some psychological cost. This assumption reflects the idea that, under a motivated cognition process (Kunda, 1987, 1990), individuals may use sophisticated mental strategies such as manipulating their own memory (Bénabou, 2015; Bénabou and Tirole, 2016)[9], avoiding freely available information (Golman, Hagmann, and Loewenstein, 2017) or creating elaborate narratives

[8]See Krizan and Windschitl (2009) for a more detailed discussion on the differences between wishful thinking and motivated reasoning.

[9]For experimental evidence on memory manipulation see, e.g., Saucet and Villeval (2019), Carlson, Maréchal, Oud, Fehr, and Crockett (2020) and Chew, Huang, and Zhao (2020).

supporting their bad choices or inaccurate claims to justify their preferred beliefs.[10] Our assumptions on the cost function captures, in "reduced form", the fact that implementing such mental strategies comes at a cost when desired beliefs deviate from from the Bayesian rational ones. In contrast, Brunnermeier and Parker (2005) model the cost of erroneous beliefs as the instrumental loss associated with the inaccurate choices induced by such beliefs. It is worth noting that Coutts (2019) provides experimental evidence in favor of the psychological rather than instrumental costs associated with belief distortion.

2.2. Model

States and prior belief. A state of the world θ is drawn by Nature from a state space Θ according to a prior distribution $\mu_0 \in \text{int}(\Delta(\Theta))$.[11] Receiver (he) and Sender (she) do not observe the state ex-ante but its prior distribution is common knowledge.

Actions and payoffs. Receiver chooses an action a from a compact space A with at least two actions. His material payoff is given by $u(a, \theta)$.[12] Receiver's choice affects Sender's payoff, which is given by $v(a)$. Before Receiver takes his action, Sender can commit to any signal structure (σ, S) given by an endogenously chosen set of signal realizations S and a stochastic mapping $\sigma\colon \Theta \to \Delta(S)$ associating any realized state θ to a conditional distribution $\sigma(\theta)$ over S.

Receiver's behavior. For any belief $\eta \in \Delta(\Theta)$, Receiver's optimal action correspondance is given by

$$A(\eta) = \arg\max_{a \in A} \int_{\Theta} u(a, \theta)\, \eta(\mathrm{d}\theta).$$

Without loss of generality, we assume that no action is dominated, i.e., for any action $a \in A$ there always exists some belief η such that $a \in A(\eta)$. When the set

[10] One can relate this possible microfoundation of the belief distortion cost to the literature on lying costs (Abeler, Becker, and Falk, 2014; Abeler, Nosenzo, and Raymond, 2019) since, when Receiver is distorting away his subjective belief from the rational Bayesian beliefs, he is essentially lying to himself. We thank Emeric Henry for suggesting us this interpretation of the cost function.

[11] In what follows, for any nonempty Polish space X, we denote $\Delta(X)$ the set of Borel probability measures over the measure space $(X, \mathcal{B}(X))$. We always endow $\Delta(X)$ with the weak*-topology. If the support of a measure $\mu \in \Delta(X)$ is finite we adopt the shorthand notation $\mu(\{x\}) = \mu(x)$ for any $x \in \text{supp}(\mu)$.

[12] We assume the map $u(a, \cdot)\colon \Theta \to \mathbb{R}$ to be Borel measurable, continuous and bounded for any $a \in A$.

$A(\eta)$ has more than one element we break the tie in favor of Sender. That is, for any belief η, the action played by Receiver in equilibrium is given by a selection $a(\eta) \in A(\eta)$ which maximizes Sender's expected payoff.[13]

Receiver's beliefs. After observing any signal realization $s \in S$, a Bayesian decision-maker's belief is given by

$$\mu(\tilde{\Theta} \mid s) = \frac{\int_{\tilde{\Theta}} \sigma(s \mid \theta)\, \mu_0(\mathrm{d}\theta)}{\int_{\Theta} \sigma(s \mid \theta)\, \mu_0(\mathrm{d}\theta)},$$

for any Borel set $\tilde{\Theta} \subseteq \Theta$.

In contrast, we assume that, when forming beliefs, Receiver trades-off the psychological benefit against the psychological cost of holding possibly non-Bayesian beliefs. The psychological benefit of Receiver under a certain belief η is given by his *anticipated material payoff*

$$U(\eta) = \int_{\Theta} u(a(\eta), \theta)\, \eta(\mathrm{d}\theta).$$

However, holding belief η when the Bayesian belief generated by some signal is μ comes at a psychological cost $C(\eta, \mu)$ for Receiver. We assume that this cost is given by the Kullback-Leibler divergence between η and μ, formally defined by

$$C(\eta, \mu) = \int_{\Theta} \frac{\mathrm{d}\eta}{\mathrm{d}\mu}(\theta) \ln\left(\frac{\mathrm{d}\eta}{\mathrm{d}\mu}(\theta)\right) \mu(\mathrm{d}\theta),$$

for any $\eta, \mu \in \Delta(\Theta)$, where $\mathrm{d}\eta/\mathrm{d}\mu$ is the Radon-Nikodym derivative of η with respect to μ, defined whenever η is absolutely continuous with respect to μ. This assumption is made for tractability but does not qualitatively affect our main results.[14] Accordingly, we define Receiver's *psychological payoff* as

$$\Psi(\eta, \mu) = U(\eta) - \frac{1}{\rho} C(\eta, \mu),$$

for any $\eta, \mu \in \Delta(\Theta)$, where $\rho \in \mathbb{R}_+^*$ parametrizes the extent of Receiver's *wish-*

[13]There might be more than one such selection if there exists some $\eta \in \Delta(\Theta)$ at which Sender is indifferent between some actions in $A(\eta)$. In that case, we pick arbitrarily one of those.

[14]We show that our results on Receiver's equilibrium beliefs and behavior continue to hold when the psychological cost functions belongs to a more general class of statistical divergences in appendix B.1.

fulness. Receiver's belief η must maximize his psychological payoff given any Bayesian belief μ. Therefore, it must belong to the optimal beliefs correspondence

$$B(\mu) = \arg\max_{\eta \in \Delta(\Theta)} \Psi(\eta, \mu),$$

for any $\mu \in \Delta(\Theta)$, and Receiver's psychological payoff when he holds a belief $\eta \in B(\mu)$ is

$$\Psi(\mu) = \max_{\eta \in \Delta(\Theta)} \Psi(\eta, \mu),$$

for any Bayesian posterior $\mu \in \Delta(\Theta)$.[15] We assume that when Receiver is psychologically indifferent between several beliefs in $B(\mu)$ he picks the one that maximizes Sender's expected utility. Therefore, Receiver's *equilibrium belief* is given by a selection $\eta(\mu) \in B(\mu)$ which maximizes Sender's expected payoff.[16] This tie breaking rule ensures that the Receiver's equilibrium belief is uniquely defined and simplifies the characterization of the optimal information policy.

Persuasion problem. We can equivalently think of Sender committing ex-ante to a signal structure (σ, S) or to an *information policy* $\tau \in \mathcal{T}(\mu_0)$, where

$$\mathcal{T}(\mu_0) = \left\{ \tau \in \Delta\left(\Delta(\Theta)\right) : \int_{\Delta(\Theta)} \mu(\tilde{\Theta})\, \tau(\mathrm{d}\mu) = \mu_0(\tilde{\Theta}) \text{ for any Borel set } \tilde{\Theta} \subseteq \Theta \right\},$$

is the set of Bayes-plausible distributions over posterior beliefs given the prior μ_0.

We assume *Sender knows Receiver is a wishful thinker*. Accordingly, she correctly anticipates the belief Receiver holds in equilibrium. Since Receiver's equilibrium belief characterizes how he would distort his belief away from any realized Bayesian posterior, Sender can choose the best information policy by backward induction, knowing: (i) which belief $\eta(\mu)$ Receiver holds in equilibrium after a posterior $\mu \in \mathrm{supp}(\tau)$ is realized and (ii) which action $a(\eta(\mu))$ Receiver

[15]As already noted by Bracha and Brown (2012) as well as Caplin and Leahy (2019), this optimization problem has a similar mathematical structure to the multiplier preferences developed in Hansen and Sargent (2008) and axiomatized in Strzalecki (2011). Precisely, the agent in Strzalecki (2011) solves

$$\max_{a \in A} \min_{\eta \in \Delta(\Theta)} \int_{\Theta} u(a, \theta)\, \eta(\mathrm{d}\theta) + \frac{1}{\rho} C(\eta, \mu), \tag{2.1}$$

for any given $\mu \in \Delta(\Theta)$. In that model, the parameter ρ measures the degree of confidence of the decision-maker in the belief μ or, in other words, the importance he attaches to belief misspecification. Conclusions on the belief distortion in that setting are naturally reversed with respect to our model: a receiver forming beliefs according to equation (2.1) would form overcautious beliefs. Studying how a rational Sender would persuade a Receiver concerned by robustness seems an interesting path for future research.

[16]Again, if Sender is indifferent between some beliefs we pick arbitrarily one of those.

chooses in equilibrium given the distorted belief $\eta(\mu)$. Sender's indirect payoff function is therefore given by

$$v(\mu) = v\left(a(\eta(\mu))\right)$$

for any $\mu \in \Delta(\Theta)$ and, hence, Sender's value from persuading a wishful Receiver under the prior μ_0 is

$$V(\mu_0) = \max_{\tau \in \mathcal{T}(\mu_0)} \int_{\Delta(\Theta)} v(\mu)\, \tau(\mathrm{d}\mu). \tag{2.2}$$

2.3. Receiver's wishful beliefs and behavior

In this section, we first extend Caplin and Leahy (2019) results by characterizing Receiver's equilibrium beliefs and behavior without imposing any restrictions on the action or state space.

To begin with, let Receiver's anticipated material payoff under action a and belief η be defined by

$$U_a(\eta) = \int_{\Theta} u(a, \theta)\, \eta(\mathrm{d}\theta).$$

Moreover, let

$$\eta_a(\mu) = \arg\max_{\eta \in \Delta(\Theta)} U_a(\eta) - \frac{1}{\rho} C(\eta, \mu),$$

be Receiver's belief motivated by action a under posterior μ and

$$\Psi_a(\mu) = \max_{\eta \in \Delta(\Theta)} U_a(\eta) - \frac{1}{\rho} C(\eta, \mu),$$

be Receiver's maximal psychological payoff motivated by action a under posterior μ. We identify Receiver's equilibrium belief $\eta(\mu)$ by: (i) finding the belief motivated by action a under μ, resulting in psychological payoff $\Psi_a(\mu)$, for any a and μ; (ii) finding which action it is optimal to motivate by maximizing $\Psi_a(\mu)$ with respect to a. proposition 7 characterizes $\eta_a(\mu)$ and $\Psi_a(\mu)$ in closed-form.

Proposition 7. *Receiver's maximal psychological payoff motivated by action a under the Bayesian posterior μ is given by*

$$\Psi_a(\mu) = \frac{1}{\rho} \ln \left(\int_{\Theta} \exp\left(\rho u(a, \theta)\right)\, \mu(\mathrm{d}\theta) \right), \tag{2.3}$$

and is attained uniquely at the belief

$$\eta_a(\mu)(\tilde{\Theta}) = \frac{\displaystyle\int_{\tilde{\Theta}} \exp\left(\rho u(a,\theta)\right)\ \mu(\mathrm{d}\theta)}{\displaystyle\int_{\Theta} \exp\left(\rho u(a,\theta)\right)\ \mu(\mathrm{d}\theta)}. \tag{2.4}$$

for any Borel set $\tilde{\Theta} \subseteq \Theta$.

Proof. See appendix B.1. □

Remark now that if the action a uniquely maximizes Receiver's psychological payoff under Bayesian posterior μ we have $\eta(\mu) = \eta_a(\mu)$. If, on the other hand, $\Psi_a(\mu) = \Psi_{a'}(\mu)$ at μ for some $a' \neq a$, meaning that Receiver is psychologically indifferent between two beliefs, then Sender breaks the tie. As a consequence, if $\mu \in \Delta(\Theta)$ satisfies

$$\Psi_a(\mu) > \Psi_{a'}(\mu), \tag{2.5}$$

for all $a' \neq a$, meaning that Receiver psychologically prefers action a to any other action a', then Receiver's equilibrium belief is given by

$$\eta(\mu)(\tilde{\Theta}) = \eta_a(\mu)(\tilde{\Theta}),$$

for any Borel set $\tilde{\Theta} \subseteq \Theta$. If $\mu \in \Delta(\Theta)$ satisfies

$$\Psi_a(\mu) = \Psi_{a'}(\mu),$$

for some $a' \neq a$, meaning that Receiver is psychologically indifferent between some actions a' and a, then Sender picks her preferred belief given by

$$\eta(\mu)(\tilde{\Theta}) = \eta_{a^*}(\mu)(\tilde{\Theta}),$$

where $a^* \in \arg\max_{\tilde{a} \in \{a,a'\}} v(\tilde{a})$.

First, we can see from equation (2.4) that Receiver only distorts beliefs that induce actions with state-dependant payoffs, i.e., Receiver's beliefs are *stakes-dependent*. Formally, for any $a \in A$, we have $\eta_a(\mu) \neq \mu$ if, and only if, there exists $\theta \neq \theta'$ such that $u(a,\theta) \neq u(a,\theta')$. Second, Receiver forms beliefs that overweight the states associated with the highest payoff, giving rise to *overoptimism*. Formally, we always have $\eta_a(\mu)(\Theta_a) \geq \mu(\Theta_a)$ for any $a \in A$ where $\Theta_a = \arg\max_{\theta \in \Theta} u(a,\theta)$. Moreover, Receiver's belief about payoff maximizing states $\eta_a(\mu)(\Theta_a)$ grows monotonically and eventually converges to 1 as Receiver's wishfulness ρ grows

from 0 to $+\infty$.[17]

As proposition 7 shows, wishful thinking leads Receiver to hold overoptimistic beliefs. The next result shows that wishful thinking distorts Receiver's behavior accordingly.

Corollary 3. *Under his equilibrium belief, Receiver's optimal action correspondence is given by*

$$A(\eta(\mu)) = \arg\max_{a \in A} \int_{\Theta} \exp\left(\rho u(a, \theta)\right) \, \mu(\mathrm{d}\theta),$$

for any $\mu \in \Delta(\Theta)$ so Receiver's equilibrium action $a(\eta(\mu))$ corresponds to Sender's preferred selection in $A(\eta(\mu))$.

Remark that this result comes as a direct consequence of proposition 7 as, by definition, any action a is optimal under the belief motivated by action a. As already observed by Caplin and Leahy (2019), the previous result states, in essence, that a Receiver forming wishful beliefs behaves as a Bayesian agent whose preferences are distorted by the function $z \mapsto \exp(\rho z)$ for any $z \in \mathbb{R}$. Importantly, from Sender's point of view, a wishful Receiver's behavior is indistinguishable from that of a Bayesian rational agent with payoff function $\exp(\rho u(a, \theta))$. Accordingly, since the function $z \mapsto \exp(\rho z)$ is strictly convex as soon as $\rho > 0$, an agent forming wishful beliefs is less risk averse than his Bayesian self.

Corollary 3 also shows that wishful thinking materializes in the form of "motivated errors" in the sense of Exley and Kessler (2019): by choosing psychologically desirable beliefs, Receiver commits systematic errors in his decision-making, i.e., acts as if he had cognitive limitations or behavioral biases relatively to a Bayesian decision-maker.

2.4. Sender's value from persuasion

In this section, we assume that the action space of Receiver is binary, so $A = \{0, 1\}$, and that Sender wants to induce $a = 1$, so $v(a) = a$. We provide necessary and sufficient conditions on Receiver's preferences under which he would take action 1 under a greater set of beliefs than a Bayesian Receiver. This allows us to compare Sender's value from persuading a wishful rather than a Bayesian Receiver as a function of the model's primitives, that is: Receiver's preferences and wishfulness.

[17]This property comes from the fact that wishful beliefs take the form of a soft-max function. For the sake of completeness we provide a proof of this result in appendix B.2.

The restriction to a binary set of actions is with loss of generality but allows better tractability.

We start by defining the two following sets of beliefs:

$$\Delta_a^B = \{\mu \in \Delta(\Theta) : a \in A(\mu)\},$$

and

$$\Delta_a^W = \{\mu \in \Delta(\Theta) : a \in A(\eta(\mu))\},$$

for any $a \in A$. The set Δ_a^B (resp. Δ_a^W) is the subset of posterior beliefs supporting an action a as optimal for a Bayesian (resp. wishful) Receiver. We say that an action is *favored* by a wishful receiver if that action is supported as optimal on a strictly larger set of posterior beliefs by a wishful Receiver compared to a Bayesian.

Definition 6 (Favored action)**.** *An action $a \in A$ is favored by a wishful Receiver if $\Delta_a^B \subset \Delta_a^W$.*

Assume for now on that $\Theta = \{\underline{\theta}, \overline{\theta}\}$. We first characterize when a wishful Receiver favors action $a = 1$ when the state space is binary and show afterwards that our results extend to any finite state space. Let us denote $u(a, \underline{\theta}) = \underline{u}_a$ and $u(a, \overline{\theta}) = \overline{u}_a$ for any $(a, \theta) \in A \times \Theta$. Assume that Receiver wants to "match the state," such that $\overline{u}_1, \underline{u}_0 > \overline{u}_0, \underline{u}_1$. Define the *payoff variability under action* 0 by $u_0 = \underline{u}_0 - \overline{u}_1$, the *payoff variability under action* 0 by $u_1 = \overline{u}_1 - \underline{u}_1$ and the indicator of the *highest achievable payoff* by $u_{\max} = \underline{u}_0 - \overline{u}_1$. With a small abuse of notation, denote $\eta = \eta(\overline{\theta})$ and $\mu = \mu(\overline{\theta})$.

By corollary 3, comparing how a wishful Receiver behaves compared to a Bayesian one is equivalent to comparing the behavior of two Bayesian receivers with respective payoff functions $\exp(\rho u(a, \theta))$ and $u(a, \theta)$. Thus, denote μ^B (resp. $\mu^W(\rho)$) the belief at which a Receiver with preferences $u(a, \theta)$ (resp. $\exp(\rho u(a, \theta))$) is indifferent between the two actions. Those beliefs are respectively equal to

$$\mu^B = \frac{\underline{u}_0 - \underline{u}_1}{\underline{u}_0 - \underline{u}_1 + \overline{u}_1 - \overline{u}_0}$$

and

$$\mu^W(\rho) = \frac{\exp(\rho \underline{u}_0) - \exp(\rho \underline{u}_1)}{\exp(\rho \underline{u}_0) - \exp(\rho \underline{u}_1) + \exp(\rho \overline{u}_1) - \exp(\rho \overline{u}_0)}.$$

With only two states, a wishful Receiver favors action $a = 1$ if and only if $\mu^W < \mu^B$, since whenever that condition is satisfied a wishful Receiver takes action $a = 1$ under a larger set of beliefs than a Bayesian. Next proposition characterizes when this is the case.

Lemma 11. *Action $a = 1$ is favored by a wishful Receiver if, and only if:*

(i) $u_{\max} \leq 0$ *and* $u_0 < u_1$*, or;*

(ii) $u_{\max} < 0$*,* $u_0 > u_1$ *and* $\rho > \overline{\rho}$*, or;*

(iii) $u_{\max} > 0$*,* $u_0 < u_1$ *and* $\rho < \overline{\rho}$*.*

where $\overline{\rho}$ *is a strictly positive threshold such that*

$$\mu^W(\overline{\rho}) = \mu^B.$$

Proof. See appendix B.3. □

Two key aspects of Receiver's material payoff thus determine which action he favors: *the highest achievable payoff* as well as *the payoff variability* for both actions. It is easy to grasp the importance of the highest payoff. Since the wishful thinker always distorts his beliefs in the direction of the most favorable outcome, in the limit, when there is no cost of distorting the Bayesian belief, Receiver would fully delude himself and always play the action that potentially yields such a payoff. The payoff variability u_a, on the other hand, is precisely Receiver's marginal psychological benefit from distorting his belief under action a. Hence, the higher the payoff variability associated with action a, the more the uncertainty about θ is relevant when such action is played and the bigger the marginal gain in anticipatory payoff the wishful thinker would get from distorting beliefs.

lemma 11 states that if an action a has both the highest payoff $\underline{u}_0$ or $\overline{u}_1$ and the greatest payoff variability u_a among all actions $a \in A$, it is always favored. If an action has either the highest payoff or the greatest payoff variability, then the wishfulness parameter ρ defines whether or not it is favored: for high wishfulness the action with the highest payoff is favored, whereas for low wishfulness it is the action with the greatest payoff variability that is favored. The intuition is the following: for sufficiently high values of Receiver's wishfulness, Receiver can afford stronger overoptimism about the most desired outcome, thus favoring the action that potentially yields this outcome despite such action not being associated with the highest marginal psychological benefit. In contrast, for sufficiently low values of ρ, Receiver cannot afford too much overoptimism about the most desired outcome. Hence, he prefers to distort beliefs at the margin that yields the highest marginal psychological benefit, such that the action associated with the highest payoff variability is favored.

The next proposition extends lemma 11 to an arbitrary finite number of states.

Proposition 8. *Assume Θ is a finite set with more than two elements. Receiver favors action $a = 1$ if, and only if, for any pair of states $\theta, \theta' \in \Theta$, Receiver's material payoffs associated with those states and his wishfulness parameter ρ satisfy one of the conditions (i), (ii) or (iii) in lemma 11.*

Proof. See appendix B.4. $\square$

Proposition 8 can easily be visualized graphically in an example with three states. Assume $\Theta = \{0, 1, 2\}$ and denote $\mu^B_{\theta,\theta'}$ (resp. $\mu^W_{\theta,\theta'}$) the belief making a Bayesian (resp. wishful) Receiver indifferent between actions $a = 0$ and $a = 1$ when $\mu(\theta), \mu(\theta') > 0$ but $\mu(\theta'') = 0$ for any $\theta, \theta', \theta'' \in \Theta$. In figure 2.1 we illustrate how Δ^W_1 compares to Δ^B_1 when Receiver's payoff function is given by:

$u(a, \theta)$	$\theta = 0$	$\theta = 1$	$\theta = 2$
$a = 0$	2	3	−1
$a = 1$	1	0	4

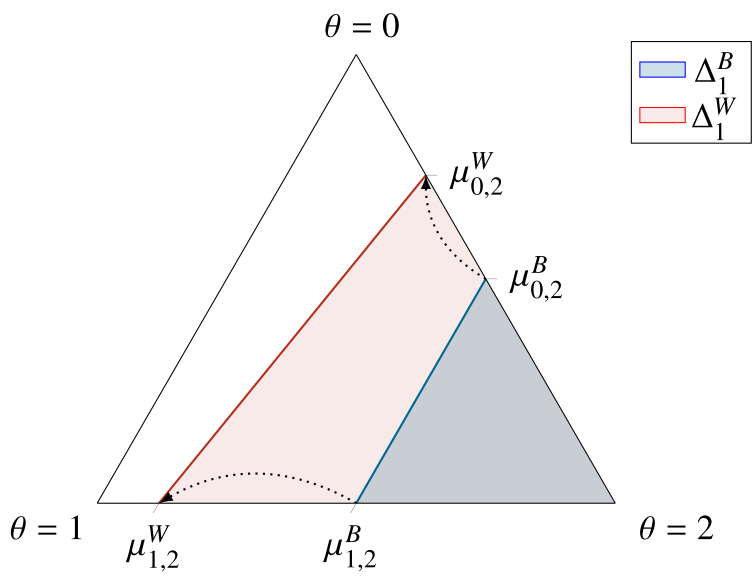


Figure 2.1: Comparison of supporting sets of beliefs. In blue, the set of Bayesian posteriors supporting action $a = 1$ for a Bayesian Receiver. In red, the set of Bayesian posteriors supporting action $a = 1$ for a wishful Receiver.

Notice that for the two pairs of states $(0, 2)$ and $(1, 2)$, the associated payoffs satisfy property (i) in lemma 11. That is, action $a = 1$ is associated with the highest payoff $u(1, 2) = 4$ as well as the highest payoff variability $u(1, 2) - u(0, 2) = 5$, under both pair of states. As a consequence, lemma 11 applies whenever focusing on those two pairs of states letting the other one being assigned probability zero. Then, we have $\mu^W_{0,2} > \mu^B_{0,2}$ and $\mu^W_{1,2} > \mu^B_{1,2}$. Remark now, that $\Delta^B_1 = \text{co}(\{\mu^B_{0,2}, \mu^B_{1,2}, \delta_2\})$

and $\Delta_1^W = \mathrm{co}(\{\mu_{0,2}^W, \mu_{1,2}^W, \delta_2\})$, where δ_θ denotes the Dirac distribution on state $\theta \in \Theta$. Consequently, $\Delta_1^B \subset \Delta_1^W$ so action $a = 1$ is favored by Receiver. If one of the conditions highlighted in lemma 11 were not satisfied for at least one of the pairs of states $(0, 2)$ or $(1, 2)$ then one of the thresholds $\mu_{\theta,\theta'}^W$ would be less or equal than $\mu_{\theta,\theta'}^B$ in which case Δ_1^W would not be a superset of Δ_1^B anymore.

Let us now turn our attention to the following questions: when is Sender better-off facing a wishful Receiver compared to a Bayesian and how does the (Blackwell) informativeness of Sender's optimal policy compare when persuading a wishful or a Bayesian Receiver? Remember that Sender chooses an information policy $\tau \in \Delta(\Delta(\Theta))$ maximizing

$$\int_{\Delta(\Theta)} v(\mu)\, \tau(\mathrm{d}\mu),$$

where

$$v(\mu) = \begin{cases} 1 & \text{if} \quad \mu \in \Delta_1^W \\ 0 & \text{otherwise} \end{cases},$$

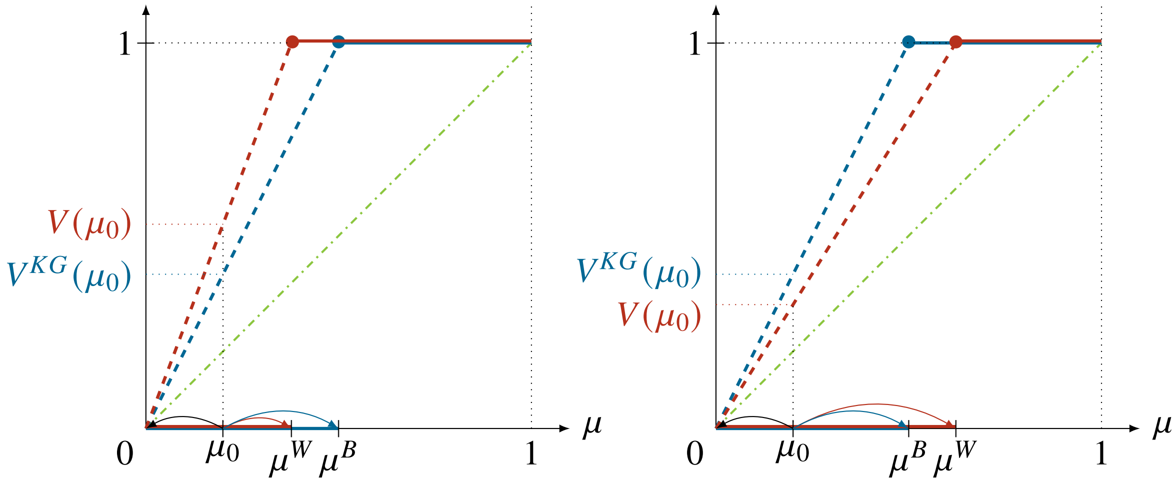
subject to the Bayes plausibility constraint

$$\int_{\Delta(\Theta)} \mu\, \tau(\mathrm{d}\mu) = \mu_0.$$

In the binary state case, it means that the threshold belief μ^W corresponds to the *smallest Bayesian posterior Sender needs to induce to persuade a wishful Receiver to take action* $a = 1$. Therefore, lemma 11 and proposition 8 have immediate consequences for Sender.

Corollary 4. *Let Θ be an arbitrary finite space with at least two elements. Then, Sender always achieves a weakly higher payoff when interacting with a wishful Receiver compared to a Bayesian for any prior $\mu_0 \in\,]0,1[$ if, and only if, for any pair of states $\theta, \theta' \in \Theta$, Receiver's material payoffs associated with those states and his wishfulness parameter ρ satisfy one of the conditions (i), (ii) or (iii) in lemma 11. Moreover, when the state space is binary, Sender's optimal information policy is always weakly less (Blackwell) informative than in the Bayesian case.*

To illustrate corollary 4 we represent in figure 2.2 the concavifications of Sender's indirect utility when Receiver is wishful or Bayesian in two different cases. The case corresponding to lemma 11 is represented in figure 2.2a. Sender is always better-off persuading a wishful compared to a Bayesian receiver as $V(\mu_0) \geq V^{KG}(\mu_0)$ for any $\mu_0 \in\,]0,1[$. On the other hand, if Receiver's preferences or wishfulness do not



(a) At least one property in lemma 11 is satisfied.

(b) No property in lemma 11 is satisfied.

Figure 2.2: Expected payoffs under optimal information policies. Red curves: expected payoffs under wishful thinking. Blue curves: expected payoffs when Receiver is Bayesian. Dashed-dotted green lines: expected payoffs under a fully revealing experiment.

satisfy any of the properties in lemma 11, then Sender is weakly worse-off under any prior. This case is represented on figure 2.2b.

When Sender wants to induce an action that is (resp. is not) favored by a wishful Receiver, persuasion is always “easier” (resp. “harder”) for Sender in the following sense: Sender needs a strictly less (resp. strictly more) Blackwell informative policy than KG to persuade Receiver to take his preferred action. Equivalently, if experiments were costly to produce, as in Gentzkow and Kamenica (2014), then Sender would always need to consume less (resp. more) resources to persuade a wishful Receiver to take his preferred action than a Bayesian. The hypothesis of a binary state space facilitates the comparisons between the Bayesian-optimal and the wishful-optimal information policies as it ensures that the Bayesian-optimal and the wishful-optimal information policies are Blackwell comparable. Although the informativeness comparisons in corollary 4 do not necessarily extend when the state space contains more than two elements, Sender’s welfare comparisons, in contrast, still hold under any arbitrary finite state space. We compare in figure 2.3 Sender’s optimal information policies when Receiver is Bayesian and wishful, with the same payoff function as in figure 2.1. When the state space is finite, a policy $\tau \in \mathcal{T}(\mu_0)$ such that all elements in $\text{supp}(\tau)$ are affinely independent is (weakly) more Blackwell-informative than a policy $\tau' \in \mathcal{T}(\mu_0)$ if, and only if, and $\text{supp}(\tau') \subset \text{co}(\text{supp}(\tau))$ (see Lipnowski, Mathevet,

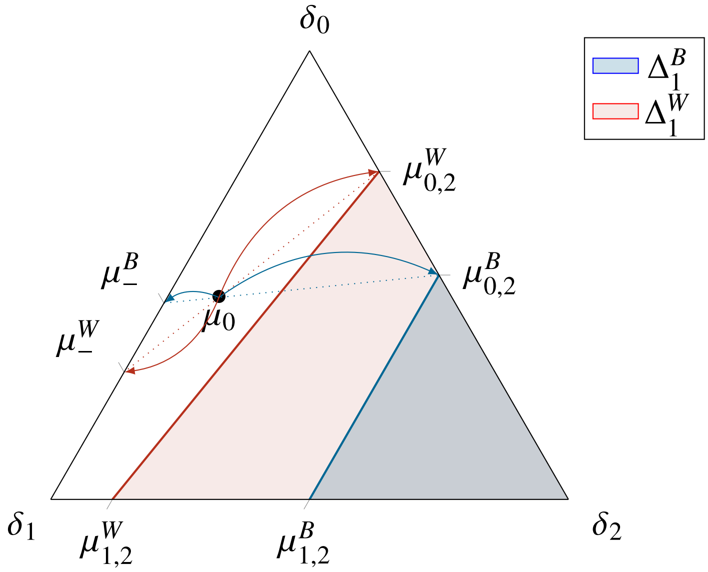


Figure 2.3: The Bayesian-optimal policy τ^B (in blue) vs. the wishful-optimal policy τ^W (in red) with respective supports $\{\mu_-^B, \mu_{0,2}^B\}$ and $\{\mu_-^W, \mu_{0,2}^W\}$.

and Wei, 2020, Lemma 2). The support of the Bayesian-optimal policy τ^B (resp. wishful-optimal policy τ^W) is $\{\mu_-^B, \mu_{0,2}^B\}$ (resp. $\{\mu_-^W, \mu_{0,2}^W\}$). Hence, $\text{co}(\text{supp}(\tau^W)) = \{\mu \in \Delta(\Theta) : \exists t \in [0,1], \mu = t\mu_-^W + (1-t)\mu_{0,2}^W\}$. It is visible on figure 2.3 that $\{\mu_-^B, \mu_{0,2}^B\} \not\subset \text{co}(\text{supp}(\tau^W))$. Hence, τ^B and τ^W are not Blackwell comparable. However, since Sender is interested in inducing action $a = 1$ and Receiver's favors that action, Sender's expected payoff is higher for any prior when Receiver is wishful.

2.5. Applications

In this section, we expose in three applications that corollary 4 might have important economic consequences.

2.5.1. Information provision and preventive health care

A public health agency (Sender) informs an individual (Receiver) about the prevalence of a certain disease. Receiver forms beliefs about the infection risk, which can be either high or low: $0 < \underline{\theta} < \overline{\theta} < 1$. The probability of contracting that illness also depends on whether the individual adopts a preventive treatment or not, where $a = 1$ designates adoption. Investment in the treatment entails a cost $c > 0$ to Receiver.[18] Moreover, let us assume that the effectiveness of the treatment,

[18] One might interpret that cost to be the price of the treatment or the either material or psychological cost from undertaking medical procedures.

i.e., the probability that the treatment works, is $\alpha \in [0, 1]$ so that the probability of falling ill, conditional on adoption, is $(1-\alpha)\theta$. The payoff from staying healthy is normalized to 0 whereas the payoff from being infected equals $-\varsigma < 0$ where ς is the severity of the disease. Receiver's payoff function is

$$u(a, \theta) = (1-a)(-\varsigma\theta) + a(-(1-\alpha)\theta\varsigma - c)$$

for any $(a, \theta) \in A \times \Theta$. We assume that $\varsigma\alpha\underline{\theta} < c < \varsigma\alpha\overline{\theta}$ so Receiver faces a trade-off: he would prefer not to invest if he was sure the probability of infection was low and, conversely, would prefer to invest in the treatment if he was sure the risk of infection is high. Also remark that Receiver always expects to experience a negative payoff, as $u(a, \theta) < 0$ for any $(a, \theta) \in A \times \Theta$.

The public health agency wants to maximize the probability of individuals adopting the preventive treatment.[19] The agency informs individuals about the prevalence of the disease by designing and committing to a Bayes-plausible information policy τ. A Bayesian Receiver would be indifferent between adopting or not the treatment at belief

$$\mu^B = \frac{c - \alpha\underline{\theta}\varsigma}{\alpha(\overline{\theta} - \underline{\theta})\varsigma}.$$

In contrast, by proposition 7 and corollary 3, the equilibrium beliefs and behavior of a wishful Receiver are given by

$$\eta(\mu) = \begin{cases} \dfrac{\mu}{\mu + (1-\mu)\exp(\rho\varsigma(\overline{\theta} - \underline{\theta}))} & \text{if} \quad \mu < \mu^W \\[2ex] \dfrac{\mu \exp(-\rho(1-\alpha)\varsigma(\overline{\theta} - \underline{\theta}))}{\mu \exp(-\rho(1-\alpha)\varsigma(\overline{\theta} - \underline{\theta})) + (1-\mu)} & \text{if} \quad \mu \geq \mu^W \end{cases},$$

and

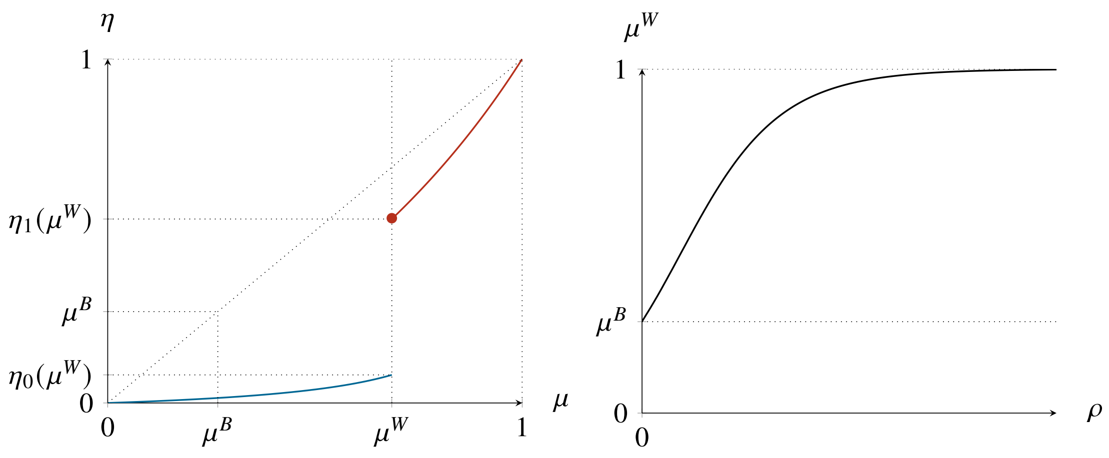
$$a(\eta(\mu)) = \mathbb{1}\left\{\mu \geq \mu^W\right\}$$

[19]Maximizing the probability of adoption is a sensible objective since most infections cause negative externalities due to their transmission through social interactions. Therefore, a benevolent planner who wants to reduce the likelihood of transmission of an infection would do well to maximize the rate of adoption of the preventive treatment (for example, maximize condom distribution to control AIDS transmission, maximize injection of vaccines to control viral infections, or maximize mask use to control the spread of airborne diseases).

for any posterior belief $\mu \in [0, 1]$, where

$$\mu^W = \frac{\exp(-\rho\underline{\theta}\varsigma) - \exp(\rho(-(1-\alpha)\underline{\theta}\varsigma - c))}{\exp(-\rho\varsigma\underline{\theta}) - \exp(\rho(-(1-\alpha)\underline{\theta}\varsigma - c)) + \exp(\rho(-(1-\alpha)\overline{\theta}\varsigma - c)) - \exp(-\rho\overline{\theta}\varsigma)}.$$

We illustrate the belief distortion of Receiver in figure 2.4a. Receiver is always



(a) equilibrium belief $\eta(\mu)$ as a function of μ.

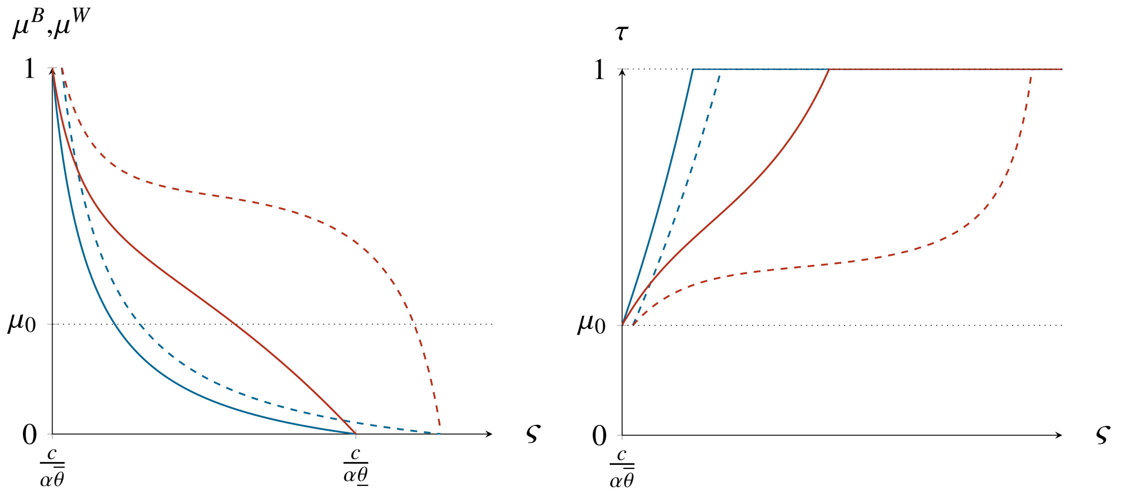
(b) Behavioral threshold μ^W as a function of ρ.

Figure 2.4: The belief correspondence for $\varsigma = 2$, $c = 0.5$, $\alpha = 0.8$, $\underline{\theta} = 0.1$, $\overline{\theta} = 0.9$ and $\rho = 2$. Receiver is always overoptimistic concerning his health risk for any induced posterior, except at $\mu = 0$ or $\mu = 1$. Moreover, the belief threshold μ^W as a function of ρ is strictly increasing and admits μ^B as a lower bound.

overoptimistic about his probability of staying healthy, as $\eta(\mu) \leq \mu$ for any $\mu \in [0, 1]$. Remark that non-adoption is associated with the highest possible payoff $-\varsigma\underline{\theta}$ as well as the highest payoff variability $\varsigma(\overline{\theta} - \underline{\theta})$. Accordingly, by lemma 11, Receiver always favors non adoption as illustrates figure 2.4b. As a result of corollary 4, Sender always needs to induce higher beliefs for Receiver to adopt the treatment than she would need if she faced a Bayesian agent, all the more so when Receiver's wishfulness ρ becomes larger. Therefore in this example, overoptimism of Receiver always goes against Sender's interest.

It is interesting to see how Sender's probability of inducing the adoption of the treatment evolves with respect to the severity of the disease ς, as well as the effectiveness of the treatment α.[20] We represent on figure 2.5b the probability that Sender induces adoption of the treatment under the optimal information policy as a function of ς. Notice that the probability of inducing adoption is less sensitive to the severity of the disease, i.e., becomes "flatter," when facing a wishful Receiver

[20] This probability is pinned down by the Bayes-plausibility constraint and equal to $\tau^{KG} = \mu_0/\mu^B$ in the Bayesian case and $\tau = \mu_0/\mu^W$ in the wishful case.



(a) Behavioral thresholds μ^B (in blue) and μ^W (in red) as functions of severity ς.

(b) Probability τ of inducing treatment adoption as a function of severity ς.

Figure 2.5: Red (resp. blue) curves correspond to wishful (resp. Bayesian) Receiver. We set parameters to $c = 0.5$, $\alpha = 0.8$, $\underline{\theta} = 0.1$, $\overline{\theta} = 0.9$ and $\rho = 2$. Full lines correspond to the case where $\alpha = 1$ whereas dashed curves correspond to $\alpha = 0.8$.

compared to the Bayesian when the treatment becomes less effective. The intuition is the following: when the treatment is fully effective, i.e., $\alpha = 1$, Receiver's payoff in case he invests in the treatment becomes state independent. Therefore, he does not have any incentive to distort beliefs when taking action $a = 1$. As a result, μ^W decreases and Receiver holds perfectly Bayesian beliefs when $\mu \geq \mu^W$. However, whenever there is uncertainty about the treatment efficacy, i.e., $\alpha < 1$, uncertainty about infection risk matters and gives room to belief distortion even when taking the treatment. Decreasing α increases the anticipated anxiety of Receiver leading to more optimistically biased beliefs, a higher μ^W and, in turn, complicates persuasion for Sender for any severity s. Remark on figure 2.5b that τ decreases sharply with α for a fixed s. In fact, one could show that as α decreases, τ becomes closer and closer to μ_0 for any ς, meaning that the agency cannot achieve a substantially higher payoff than under full disclosure.[21]

In the next subsection we extend out framework to the case of a continuous state space and linear preferences. We show that results in the finite state space case extend to this setting. We also highlight why we might expect persuasion to

[21] One additional implication of this result is the following. Assume the true treatment efficacy is α but Receiver perceives the efficacy to be $\hat{\alpha} < \alpha$ (e.g. because Receiver adheres to anti-vaccines movements or generally mistrusts the pharmaceutical industry). In that case, the doubts expressed by Receiver about the treatment efficacy makes him even more anxious which, in turn, makes belief distortion stronger and, thus, downplays the effectiveness of the agency's information policy whatever is the severity of the disease.

be more effective in the context of risky investment decisions.

2.5.2. Persuading a wishful investor

A financial broker (Sender) designs reports about the return of some risky financial product to inform a potential client (Receiver). The return of the product is $\theta \in \Theta = [\underline{\theta}, \overline{\theta}]$, where $\underline{\theta} < 0 < \overline{\theta}$. Returns are distributed according to the prior distribution μ_0. Let F be the cumulative distribution function associated with μ_0 and let us assume that μ_0 admits a continuous and strictly positive density function f over $[\underline{\theta}, \overline{\theta}]$. Receiver has some saved up money he is willing to invest and chooses action $a \in A = \{0, 1\}$, where $a = 0$ represents the choice of non-investing in which case Receiver's payoff is 0 and $a = 1$ represents investing, in which case Receiver's payoff is the realized return θ. The broker is remunerated on the basis of a flat fee $v > 0$ that is independent of the true product's profitability. Hence, Receiver's payoff is $u(a, \theta) = a\theta$ while Sender's payoff is $v(a, \theta) = va$ for any $(a, \theta) \in A \times \Theta$.

Receiver forms motivated beliefs about the return of the financial product. By proposition 7 his equilibrium beliefs are given by

$$\eta(\mu)(\tilde{\Theta}) = \begin{cases} \mu(\tilde{\Theta}) & \text{if } \displaystyle\int_{\Theta} \exp(\rho\theta)\, \mu(\mathrm{d}\theta) < 1 \\ \dfrac{\displaystyle\int_{\tilde{\Theta}} \exp(\rho\theta)\, \mu(\mathrm{d}\theta)}{\displaystyle\int_{\Theta} \exp(\rho\theta)\, \mu(\mathrm{d}\theta)} & \text{if } \displaystyle\int_{\Theta} \exp(\rho\theta)\, \mu(\mathrm{d}\theta) \geq 1 \end{cases},$$

for any $\mu \in \Delta(\Theta)$ and any Borel set $\tilde{\Theta} \subseteq \Theta$, and, by corollary 3, his equilibrium behavior is given by

$$a(\eta(\mu)) = \mathbb{1}\left\{\int_{\Theta} \exp(\rho\theta)\, \mu(\mathrm{d}\theta) \geq 1\right\}.$$

Therefore, Sender's indirect utility is equal to

$$v(\mu) = v\mathbb{1}\left\{\int_{\Theta} \exp(\rho\theta)\, \mu(\mathrm{d}\theta) \geq 1\right\}.$$

for any $\mu \in \Delta(\Theta)$. To make the problem interesting, we assume that neither a Bayesian nor a wishful Receiver would take action $a = 0$ under the prior. That is, $\hat{m} = \int_{\underline{\theta}}^{\overline{\theta}} \theta\mu_0(\mathrm{d}\theta) < 0$ and $\hat{x} = \int_{\underline{\theta}}^{\overline{\theta}} \exp(\rho\theta)\mu_0(\mathrm{d}\theta) < 1$.[22]

[22] It is in fact always true that $\hat{m} < 0$ when $\hat{x} < 1$. Hence, assuming $\hat{m} < 0$ additionally to $\hat{x} < 1$

Under these assumptions, remark that a signal structure σ that induces a distribution τ over posterior beliefs μ matters for Receiver and Sender only through the *distribution of exponential moments* $x = \int_\Theta \exp(\rho\theta)\,\mu(\mathrm{d}\theta)$ *it induces*. Let X be the space of such moments, that is, $X = \mathrm{co}(\exp(\rho\Theta))$, where $\exp(\rho\Theta)$ is the graph of the function $\theta \mapsto \exp(\rho\theta)$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$. That is, $X = [\underline{x}, \overline{x}]$ where $\underline{x} = \exp(\rho\underline{\theta})$ and $\overline{x} = \exp(\rho\overline{\theta})$. Let G be the prior cumulative distribution function over the random variable $\exp(\rho\theta)$ induced by F, that is

$$G(x) = F\left(\frac{\ln(x)}{\rho}\right),$$

for any $x \in [\underline{x}, \overline{x}]$. By standard arguments (Gentzkow and Kamenica, 2016), the problem of finding an optimal signal structure σ reduces to finding a cumulative distribution function H that maximizes

$$\int_{\underline{x}}^{\overline{x}} v(x)\,\mathrm{d}H(x)$$

subject to

$$\int_{\underline{x}}^{z} H(x)\,\mathrm{d}x \leq \int_{\underline{x}}^{z} G(x)\,\mathrm{d}x$$

for every $z \in [\underline{x}, \overline{x}]$. The solution to such a problem is well-known and can be found either using techniques from optimization under stochastic dominance constraints (Gentzkow and Kamenica, 2016; Ivanov, 2020; Kleiner, Moldovanu, and Strack, 2021) or linear programming (Kolotilin, 2018; Dworczak and Martini, 2019; Dizdar and Kováč, 2020). In our context, the optimal signal is a binary partition of the state space. That is, the broker reveals whether the return is above or below some threshold state.

Proposition 9. *There exists a unique* $\theta^W \in [\underline{\theta}, \overline{\theta}]$ *verifying*

$$\frac{1}{1 - F(\theta^W)} \int_{\theta^W}^{\overline{\theta}} \exp(\rho\theta) f(\theta)\,\mathrm{d}\theta = 1$$

and such that Sender pools all states $\theta \in [\theta^W, \overline{\theta}]$ *under the same signal* $s = 1$*, i.e.,* $\sigma(1 \mid \theta) = 1$ *for all* $\theta \in [\theta^W, \overline{\theta}]$*, and similarly pools all states* $\theta \in [\underline{\theta}, \theta^W]$ *under the same signal* $s = 0$*. Hence, the probability of inducing action* $a = 1$ *for Sender*

is without loss.

is equal to

$$\int_{\theta^W}^{\overline{\theta}} \sigma(1 \mid \theta) f(\theta) \, \mathrm{d}\theta = 1 - F(\theta^W).$$

Proof. See Ivanov (2020), Section 3. □

It is optimal for Sender to partition the state space at the threshold state making Receiver indifferent between investing or not at the prior. Such an information policy can intuitively be seen as the investment recommendation rule which maximizes the probability that Receiver invests given the prior distribution of returns F.

Using the exact same arguments as above, one can deduce that the probability of inducing action $a = 1$ when Receiver is Bayesian is given by $1 - F(\theta^B)$ where θ^B is the unique threshold verifying the equation

$$\frac{1}{1 - F(\theta^B)} \int_{\theta^B}^{\overline{\theta}} \theta f(\theta) \, \mathrm{d}\theta = 0.$$

Therefore, Sender is more effective at persuading a wishful Receiver if and only if $\theta^W < \theta^B$.

Proposition 10. *It is always true that $\theta^W < \theta^B$. Hence, Sender is always more effective at persuading a wishful rather than a Bayesian investor.*

Proof. See appendix B.6. □

The above result relates to proposition 8: buying the risky product is favored by the wishful investor since it is the action that yields both the highest possible payoff and the highest payoff variability. This example thus illustrates how the results in the finite state space case naturally extend to an infinite state space setting with linear preferences. It further helps explaining the pervasiveness of persuasion efforts in financial and betting markets, illustrating why some financial consulting firms seem to specialize in advice misconduct and cater to biased consumers.

2.5.3. Public persuasion and political polarization

A Sender (e.g., a politician, a lobbyist) persuades an odd-numbered finite group of voters $N = \{1, \ldots, n\}$ (e.g., a committee or parliamentary members) to adopt a proposal $x \in X = \{0, 1\}$, where $x = 0$ corresponds to the status-quo. The state space is binary, $\Theta = \{0, 1\}$, and the audience uses only the information disclosed by Sender to vote on the proposal. Let $a^i \in A = \{0, 1\}$ be the ballot cast by voter i, where $a^i = 0$ designates voting for the status-quo. The proposal is accepted if it is

supported by a simple majority of voters. We assume Sender is only interested in the proposal being accepted, so her utility is $v(x) = x$. In contrast, any voter $i \in N$ has payoff function

$$u^i(x,\theta) = x\theta\beta^i + (1-x)(1-\theta)(1-\beta^i)$$

for any $(x,\theta) \in X \times \Theta$ where $\beta^i \in [0,1]$ parametrizes the partisan preference of voter i. That is, all voters agree that the proposal should be implemented only when $\theta = 1$, but they vary in how much they value the implementation of the proposal. We assume β^i is symmetrically distributed around $1/2$ in the population. Denote $\beta^m = 1/2$ the median voter's preference.

All voters form wishful beliefs and ρ is assumed homogeneous among the electorate. As a result, the direction as well as the magnitude of voters' belief distortion depends only on their partisan preferences β.[23] By proposition 7, voter i's belief under posterior $\mu \in [0,1]$ is given by

$$\eta(\mu,\beta^i) = \begin{cases} \dfrac{\mu}{\mu + (1-\mu)\exp(\rho(1-\beta^i))} & \text{if} \quad \mu < \mu^W(\beta^i) \\ \\ \dfrac{\mu\exp(\rho\beta^i)}{\mu\exp(\rho\beta^i) + (1-\mu)} & \text{if} \quad \mu \geq \mu^W(\beta^i) \end{cases}.$$

where

$$\mu^W(\beta^i) = \frac{\exp(\rho(1-\beta^i)) - 1}{\exp(\rho(1-\beta^i)) + \exp(\rho\beta^i) - 2}.$$

Remark that, similarly as in Alonso and Câmara (2016), since the policy space is binary and voters do not hold private information there is no room for strategic voting in our model. Hence, citizen i's voting strategy under belief $\eta(\mu,\beta^i)$ is given by

$$a(\eta(\mu,\beta^i)) = \mathbb{1}\left\{\mu \geq \mu^W(\beta^i)\right\}.$$

Due to the heterogeneity in β, there is always some level of belief polarization among wishful voters for any $\mu \in]0,1[$. Let us measure such polarization by the sum of the absolute difference between each pair of beliefs in the audience

$$\pi(\mu) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} |\eta(\mu,\beta^i) - \eta(\mu,\beta^j)| \tag{2.6}$$

[23]It has been shown in psychology (Babad, Hills, and O'Driscoll, 1992; Babad, 1995, 1997) as well as in behavioral economics (Thaler, 2020) that voters political beliefs are often motivated by their partisan orientation.

for any $\mu \in [0, 1]$.

Proposition 11. *Under Sender's optimal information policy, the signal that leads to the implementation of the proposal also generates the maximum polarization among voters.*

Proof. See appendix B.5. $\square$

To build an intuition of why this is the case, let's first note that, in our model, belief polarization and action polarization are closely related. Agents voting for the implementation of the proposal distort their beliefs upwards, whereas agents voting for the status quo distort their beliefs downwards. We can thus see that maximum belief polarization should be attained for some belief for which action polarization is maximized, that is, for some belief at which $(n+1)/2$ agents are voting one way and the remaining $(n-1)/2$ are voting another way. This is the case for any $\mu \in [\mu^W(\beta^{m-1}), \mu^W(\beta^{m+1})[$.

Due to sincere voting, the result of the election always coincides with the vote of the median voter under posterior belief μ. Accordingly, Sender's indirect utility is

$$v(\mu) = \mathbb{1}\left\{\mu \geq \mu^W(\beta^m)\right\},$$

for any $\mu \in [0, 1]$. The optimal information policy for Sender is thus supported on $\{0, \mu^W(\beta^m)\}$ whenever $\mu_0 \in]0, 1/2[$, and on $\{\mu_0\}$ whenever $\mu_0 \in]\mu^W(\beta^m), 1[$. The posterior $\mu^W(\beta^m)$, which leads to the implementation of the proposal, belongs to the interval $[\mu^W(\beta^{m-1}), \mu^W(\beta^{m+1})[$ and, as such, is in the neighbourhood of the belief that maximizes polarization for any distribution of preferences. When such distribution is symmetric around the median voter, polarization is maximized exactly at the middle point in that interval, which is $\mu^W(\beta^m)$.

We illustrate proposition 11 below in section 2.5.3 in a setup with 3 voters. Following corollary 3, wishful thinking induces voters to switch from disapproval to approval at different Bayesian posteriors $\mu^W(\beta^i)$. The optimal information policy τ for Sender is the one that maximizes the probability of the median voter voting for the approval. That is, $\text{supp}(\tau) = \{0, \mu^W(\beta^m)\}$ and $\mu^W(\beta^m) = 1/2$ is induced with probability $\tau = \mu^W(\beta^m)/\mu_0$ whenever $\mu_0 \in]0, \mu^W(\beta^2)[$ and $\text{supp}(\tau) = \{\mu_0\}$ whenever $\mu_0 \in]\mu^W(\beta^2), 1[$.

Let us now turn to polarization. First, it is quite easy to see in section 2.5.3 that

$$\pi(\mu) = 2\left(\eta(\mu, \beta^1) - \eta(\mu, \beta^3)\right)$$

for any $\mu \in [0, 1]$, as the distances to the median belief add up to $\eta(\mu, \beta^1) - \eta(\mu, \beta^3)$.

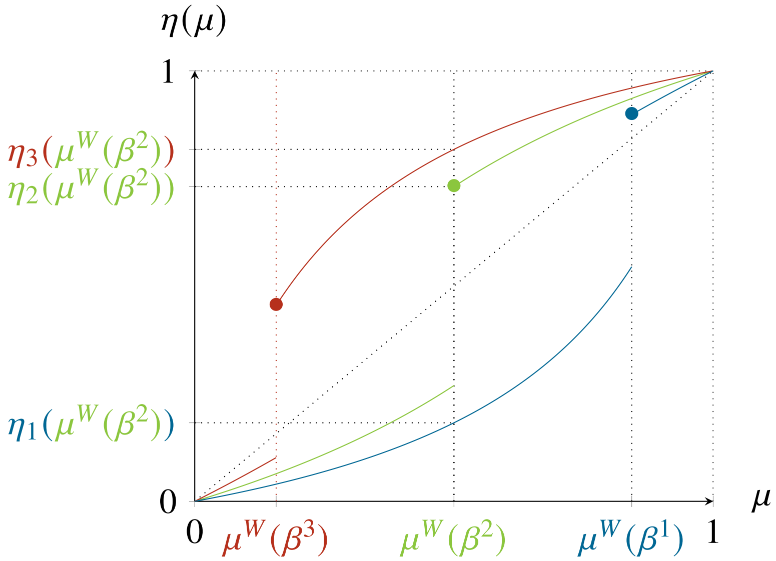


Figure 2.6: Beliefs distortions in the electorate for $\rho = 2$, $\beta_1 = 1/4$, $\beta_2 = 1/2$ and $\beta_3 = 3/4$. Polarization equals $\pi(\mu) = 2(\eta(\mu, \beta^1) - \eta(\mu, \beta^3))$ which is maximized at $\mu^W(\beta^2) = 1/2$.

Thus, it suffices to check where $\eta(\mu, \beta^1) - \eta(\mu, \beta^3)$ is maximized. Quite naturally, polarization is maximized when the posterior belief induced by Sender is in between $\mu^W(\beta^3)$ and $\mu^W(\beta^1)$. In particular, it is exactly maximized at the posterior belief $\mu^W(\beta^2) = 1/2$ which is exactly the posterior belief Sender induces to obtain the approval of the proposal under her optimal policy.

proposition 11 establishes that the intuition developed in this example is generally valid when the partisan preferences of voters are symmetrically distributed around the median. In other words, attempts by a rational sender to maximize the probability of approval induces, as an externality, maximal belief polarization among wishful voters. This result differs from the literature studying the possible heterogeneity of beliefs due to deliberate attempts at persuasion which tends to focus on polarization arising from differential access to information.[24] Our model gives an alternative mechanism to the rise of polarization, based on motivated beliefs: a sender can induce polarization involuntarily when her message is subject to motivated interpretations, and such polarization might be especially large whenever sender's strategy involves targeting an agent with a median preference.

[24]See Arieli and Babichenko (2019) for general considerations on the private persuasion of multiple receivers and see Chan, Gupta, Li, and Wang (2019) for an application to voting.

2.6. Conclusion

In this paper we study optimal persuasion in the presence of a wishful Receiver. By modeling wishful thinking as a process that optimally trades-off gains in anticipatory utility with the cost of distorting beliefs, we characterize the correspondence between wishful and Bayesian beliefs, highlighting the particularities that such belief formation process entails.

In particular, we show that wishful thinking impacts behavior, causing some actions to be favored in the sense that they are taken at a greater set of beliefs. This has important implications for the strategic design of information, as it adds some nuance on the way preferences and information determine behavior. Concretely, we show that, in the presence of wishful thinking, persuasion is more effective when it is aimed at inducing actions that are risky but can potentially yield a very large payoff and less effective when it is aimed at inducing more cautious actions. We use this model to illustrate why information disclosure seems less effective than expected at inducing preventive health behavior and more effective than expected at inducing dubious financial investments. Wishful thinking opens a channel for preferences to interfere in belief formation, raising the question of what kind of belief polarization could we observe in a population in which agents have access to the same information but vary in their preferences. We show in an application that an information designer interested in the approval of a proposal would, by optimally targeting the median voter in her choice of signal structure, induce, as an externality, maximum polarization among the electorate whenever the proposal is approved.

Some studies already investigate the effects of wishful thinking on the outcomes of strategic interactions (see, Yildiz, 2007; Banerjee, Davis, and Gondhi, 2020; Heller and Winter, 2020). Further investigation on ways in which individual preferences might impact information processing and how these may impact social phenomena such as belief polarization in non-strategic and strategic settings seem to be promising paths for future research.

Bibliography

Abeler, J., A. Becker, and A. Falk (2014): "Representative evidence on lying costs," *Journal of Public Economics*, 113, 96–104.

Abeler, J., D. Nosenzo, and C. Raymond (2019): "Preferences for Truth-Telling," *Econometrica*, 87, 1115–1153.

Akerlof, G. A. and W. T. Dickens (1982): "The Economic Consequences of Cognitive Dissonance," *American Economic Review*, 72, 307–319.

Alonso, R. and O. Câmara (2016): "Persuading Voters," *American Economic Review*, 106, 3590–3605.

Arieli, I. and Y. Babichenko (2019): "Private Bayesian persuasion," *Journal of Economic Theory*, 182, 185–217.

Babad, E. (1995): "Can Accurate Knowledge Reduce Wishful Thinking in Voters' Predictions of Election Outcomes?" *The Journal of Psychology*, 129, 285–300.

——— (1997): "Wishful thinking among voters: motivational and cognitive Iinfluences," *International Journal of Public Opinion Research*, 9, 105–125.

Babad, E., M. Hills, and M. O'Driscoll (1992): "Factors Influencing Wishful Thinking and Predictions of Election Outcomes," *Basic and Applied Social Psychology*, 13, 461–476.

Babad, E. and Y. Katz (1991): "Wishful Thinking—Against All Odds," *Journal of Applied Social Psychology*, 21, 1921–1938.

Banerjee, S., J. Davis, and N. Gondhi (2020): "Motivated Beliefs in Coordination Games," *SSRN Electronic Journal*.

Bénabou, R. (2015): "The Economics of Motivated Beliefs," *Revue d'économie politique*, 125, 665–685.

Bénabou, R. and J. Tirole (2002): "Self-Confidence and Personal Motivation," *Quarterly Journal of Economics*, 117, 871–915.

——— (2004): "Willpower and Personal Rules," *Journal of Political Economy*, 112, 848–886.

——— (2006): "Belief in a Just World and Redistributive Politics," *Quarterly Journal of Economics*, 121, 699–746.

——— (2011): "Identity, Morals, and Taboos: Beliefs as Assets *," *Quarterly Journal of Economics*, 126, 805–855.

——— (2016): "Mindful Economics: The Production, Consumption, and Value of Beliefs," *Journal of Economic Perspectives*, 30, 141–164.

Benjamin, D., A. Bodoh-Creed, and M. Rabin (2019): "Base-Rate Neglect: Foundations and Implications," .

Benjamin, D. J. (2019): "Errors in probabilistic reasoning and judgment biases," in *Handbook of Behavioral Economics: Applications and Foundations 2*, Elsevier B.V., vol. 2, chap. 2, 69–186.

Bergemann, D. and S. Morris (2016): "Information Design, Bayesian Persuasion, and Bayes Correlated Equilibrium," *American Economic Review*, 106, 586–591.

——— (2019): "Information Design: A Unified Perspective," *Journal of Economic Literature*, 57, 44–95.

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2018): "Behavioral Household Finance," in *Handbook of Behavioral Economics: Applications and Foundations 1*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, Elsevier B.V., chap. 3, 177–276.

Bracha, A. and D. J. Brown (2012): "Affective decision making: A theory of optimism bias," *Games and Economic Behavior*, 75, 67–80.

Brunnermeier, M. K. and J. A. Parker (2005): "Optimal Expectations," *American Economic Review*, 95, 1092–1118.

Caplin, A. and J. Leahy (2001): "Psychological Expected Utility Theory and Anticipatory Feelings," *Quarterly Journal of Economics*, 116, 55–79.

——— (2019): "Wishful Thinking," *NBER Working Paper Series*.

Carlson, R. W., M. A. Maréchal, B. Oud, E. Fehr, and M. J. Crockett (2020): "Motivated misremembering of selfish decisions," *Nature Communications*, 11, 2100.

Chan, J., S. Gupta, F. Li, and Y. Wang (2019): "Pivotal persuasion," *Journal of Economic Theory*, 180, 178–202.

Chandra, A., B. Handel, and J. Schwartzstein (2019): "Behavioral economics and health-care markets," in *Handbook of Behavioral Economics: Applications and Foundations 2*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, Elsevier B.V., chap. 6, 459–502.

Chew, S. H., W. Huang, and X. Zhao (2020): "Motivated False Memory," *Journal of Political Economy*, 128, 3913–3939.

Coutts, A. (2019): "Testing models of belief bias: An experiment," *Games and Economic Behavior*, 113, 549–565.

de Clippel, G. and X. Zhang (2020): "Non-Bayesian Persuasion," *Working Paper*.

Dizdar, D. and E. Kováč (2020): "A simple proof of strong duality in the linear persuasion problem," *Games and Economic Behavior*, 122, 407–412.

Dupas, P. (2011): "Health Behavior in Developing Countries," *Annual Review of Economics*, 3, 425–449.

Dworczak, P. and G. Martini (2019): "The Simple Economics of Optimal Persuasion," *Journal of Political Economy*, 127, 1993–2048.

Egan, M., G. Matvos, and A. Seru (2019): "The Market for Financial Adviser Misconduct," *Journal of Political Economy*, 127, 233–295.

Eliaz, K., R. Spiegler, and H. C. Thysen (2021a): "Persuasion with endogenous misspecified beliefs," *European Economic Review*, 134, 103712.

——— (2021b): "Strategic interpretations," *Journal of Economic Theory*, 192, 105192.

Engelmann, J., M. Lebreton, P. Schwardmann, J. J. van der Weele, and L.-A. Chang (2019): "Anticipatory Anxiety and Wishful Thinking," *SSRN Electronic Journal*.

Ettinger, D. and P. Jehiel (2010): "A Theory of Deception," *American Economic Journal: Microeconomics*, 2, 1–20.

Exley, C. and J. Kessler (2019): "Motivated Errors," *NBER Working Paper Series*.

Eyster, E. (2019): "Errors in strategic reasoning," in *Handbook of Behavioral Economics: Applications and Foundations 2*, Elsevier B.V., vol. 2, chap. 3, 187–259.

Ganguly, A. and J. Tasoff (2017): "Fantasy and Dread: The Demand for Information and the Consumption Utility of the Future," *Management Science*, 63, 4037–4060.

Gentzkow, M. and E. Kamenica (2014): "Costly Persuasion," *American Economic Review: Papers & Proceedings*, 104, 457–462.

——— (2016): "A Rothschild-Stiglitz Approach to Bayesian Persuasion," *American Economic Review: Papers & Proceedings*, 106, 597–601.

Golman, R., D. Hagmann, and G. Loewenstein (2017): "Information Avoidance," *Journal of Economic Literature*, 55, 96–135.

Golman, R., G. Loewenstein, K. O. Moene, and L. Zarri (2016): "The Preference for Belief Consonance," *Journal of Economic Perspectives*, 30, 165–188.

Hagenbach, J. and F. Koessler (2020): "Cheap talk with coarse understanding," *Games and Economic Behavior*, 124, 105–121.

Hansen, L. P. and T. J. Sargent (2008): *Robustness*, Princeton University Press.

Heger, S. A. and N. W. Papageorge (2018): "We should totally open a restaurant: How optimism and overconfidence affect beliefs," *Journal of Economic Psychology*, 67, 177–190.

Heller, Y. and E. Winter (2020): "Biased-Belief Equilibrium," *American Economic Journal: Microeconomics*, 12, 1–40.

Ivanov, M. (2020): "Optimal monotone signals in Bayesian persuasion mechanisms," *Economic Theory*.

Jiao, P. (2020): "Payoff-Based Belief Distortion," *The Economic Journal*, 130, 1416–1444.

Kamenica, E. (2019): "Bayesian Persuasion and Information Design," *Annual Review of Economics*, 11, 249–272.

Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

Kleiner, A., B. Moldovanu, and P. Strack (2021): "Extreme Points and Majorization: Economic Applications," *Econometrica*, 89, 1557–1593.

Kolotilin, A. (2018): "Optimal information disclosure: A linear programming approach," *Theoretical Economics*, 13, 607–635.

Kremer, M., G. Rao, and F. Schilbach (2019): "Behavioral development economics," in *Handbook of Behavioral Economics: Applications and Foundations 2*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, Elsevier B.V., chap. 5, 345–458.

Krizan, Z. and P. D. Windschitl (2009): "Wishful Thinking about the Future: Does Desire Impact Optimism?" *Social and Personality Psychology Compass*, 3, 227–243.

Kunda, Z. (1987): "Motivated inference: Self-serving generation and evaluation of causal theories." *Journal of Personality and Social Psychology*, 53, 636–647.

——— (1990): "The case for motivated reasoning." *Psychological Bulletin*, 108, 480–498.

Le Yaouanq, Y. (2021): "Motivated cognition in a model of voting," *Working Paper*.

Lerman, C., C. Hughes, S. J. Lemon, D. Main, C. Snyder, C. Durham, S. Narod, and H. T. Lynch (1998): "What you don't know can hurt you: adverse psychologic effects in members of BRCA1-linked and BRCA2-linked families who decline genetic testing." *Journal of Clinical Oncology*, 16, 1650–1654.

Levy, G., I. Moreno de Barreda, and R. Razin (2018): "Persuasion with Correlation Neglect," *Working Paper*.

Lipnowski, E., L. Mathevet, and D. Wei (2020): "Attention Management," *American Economic Review: Insights*, 2, 17–32.

Loewenstein, G. (1987): "Anticipation and the Valuation of Delayed Consumption," *The Economic Journal*, 97, 666–684.

Mayraz, G. (2011): "Wishful Thinking," *SSRN Electronic Journal*.

Mijović-Prelec, D. and D. Prelec (2010): "Self-deception as self-signalling: a model and experimental evidence," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365, 227–240.

Mullainathan, S., M. Noeth, and A. Schoar (2012): "The Market for Financial Advice: An Audit Study," Tech. rep., National Bureau of Economic Research, Cambridge, MA.

Mullainathan, S., J. Schwartzstein, and A. Shleifer (2008): "Coarse Thinking and Persuasion *," *Quarterly Journal of Economics*, 123, 577–619.

Oster, E., I. Shoulson, and E. R. Dorsey (2013): "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease," *American Economic Review*, 103, 804–830.

Saucet, C. and M. C. Villeval (2019): "Motivated memory in dictator games," *Games and Economic Behavior*, 117, 250–275.

Schwardmann, P. (2019): "Motivated health risk denial and preventative health care investments," *Journal of Health Economics*, 65, 78–92.

Strzalecki, T. (2011): "Axiomatic Foundations of Multiplier Preferences," *Econometrica*, 79, 47–73.

Thaler, M. (2020): "The 'Fake News' Effect: Experimentally Identifying Motivated Reasoning Using Trust in News," *SSRN Electronic Journal*.

Weinstein, N. D. (1980): "Unrealistic optimism about future life events." *Journal of Personality and Social Psychology*, 39, 806–820.

Yildiz, M. (2007): "Wishful Thinking in Strategic Environments," *Review of Economic Studies*, 74, 319–344.

3. Price Discrimination with Redistributive Concerns[1]

Abstract

Consumer data can be used to sort consumers into different market segments, allowing a monopolist to charge different prices at each segment. We study consumer-optimal segmentations with redistributive concerns, i.e., that prioritize poorer consumers. Such segmentations are efficient but may grant additional profits to the monopolist, compared to consumer-optimal segmentations with no redistributive concerns. We characterize the markets for which this is the case and provide a procedure for constructing optimal segmentations given a strong redistributive motive. For the remaining markets, we show that the optimal segmentation is surprisingly simple: it generates one segment with a discount price and one segment with the same price that would be charged if there were no segmentation.

3.1. Introduction

Consumers are continuously leaving traces of their identities on the internet, be it through social media activity, search-engine utilization, online-purchasing and so on. The vast amount of consumer data that is generated and collected has acquired the status of a highly-valued good, as it allows firms to tailor advertisements and prices to different consumers. In practice, the availability of consumer data *segments* consumers: observing that a given consumer has certain characteristics allows firms to fine-tune how they interact with people that share those characteristics. Adjusting how coarse-grained the information available about consumers is impacts

[1]This chapter is a joint work with Daniel Barreto and Alexis Ghersengorin. We thank Eduardo Perez-Richet for his guidance on this project. We also thank Matthew Elliott, Jeanne Hagenbach, Emeric Henry, Emir Kamenica, Frédéric Koessler, Shengwu Li, Franz Ostrizek, Nikhil Vellodi, Colin Stewart their valuable feedbacks and comments at various stages of the project, as well as seminar audiences at Sciences Po, Paris School of Economics, University of Konstanz, CUNEF, University of Rome "Tor Vergata," University of Barcelona, University of Amsterdam and WU Vienna for helpful discussions. All remaining errors are ours. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement 850996 – MOREV and 101001694 – IMEDMC).

how they will be segmented, what sort of digital market interactions they will have and what prices they will pay. This suggests room for regulatory oversight.

As shown by Bergemann, Brooks, and Morris (2015), consumer segmentation and price discrimination can induce a wide range of welfare outcomes. It can not only be used to increase social surplus—by creating segments with prices that allow more consumers to buy—, but can also be performed in a way that ensures that all created surplus accrues to consumers — that is, that maximizes consumer surplus. This is done by creating segments that pool together consumers with high and low willingness to pay, thus allowing higher willingness to pay consumers to benefit from lower prices. However, an important aspect of price discrimination that remains overlooked by the literature is its *distributive effect*: since different consumers pay different prices, this practice defines how surplus is distributed *across* consumers, raising questions about how it can benefit poorer consumers relative to richer ones. Indeed, if willingness to pay and wealth are positively related, segmentations that maximize total consumer surplus tend to benefit richer consumers.

In this paper we provide a normative analysis of the distributive impacts of market segmentation. Our aim is to study how this practice impacts different consumers and how it should be performed under the objective of increasing consumer welfare while prioritizing poorer consumers. Our results draw qualitative characteristics of segmentations that achieve this goal, which can be used to inform future regulation. Importantly, our analysis also shows that the prioritization of poorer consumers can be inconsistent with the maximization of total consumer surplus: raising the surplus of poorer consumers may only be possible while granting additional profits to the producer, at the expense of richer consumers.

We consider a setting in which a monopolist sells a good on a market composed of heterogeneous consumers, each of whom can consume at most one unit and is characterized by their willingness to pay for the good. A social planner can provide information about consumers' willingness to pay to the monopolist. The information provision strategy effectively divides the aggregate pool of consumers into different *segments*, each of which can be priced differently by the monopolist. The social planner's objective is to maximize a weighted sum of consumers' surplus. As in Dworczak, Kominers, and Akbarpour (2021), we consider weights that are decreasing on the consumer's willingness to pay, capturing the notion of a redistributive motive under the assumption that consumers with higher willingness to pay are on average richer than those with lower willigness to pay.

We first establish that optimal segmentations are Pareto efficient, such that

satisfying a redistributive objective does not come at the expense of social surplus. Bergemann et al. (2015) show that, in the absence of redistributive concerns, consumer-optimal segmentations do not strictly benefit the monopolist: all of the surplus created by the segmentation accrues to consumers. In contrast, we show that once redistributive preferences are considered, consumer-optimal segmentations may imply additional profits to the monopolist. This happens because increasing the surplus of poor consumers is done by pooling them with even poorer consumers, such that they can benefit from lower prices. In doing so, richer consumers become more representative in other segments, which might increase the price they pay. We characterize the set of markets for which this is the case and denote them as rent markets. For no-rent markets, on the contrary, *any* redistributive objective can be met while still maximizing total consumer surplus. In this case, our analysis selects one among the many consumer-optimal segmentations established by Bergemann et al. (2015). These insights are illustrated through a three-type example in section 3.3.

Our analysis also provides insights on how to construct optimal segmentations. We show that, in no-rent markets, consumer-optimal segmentations with redistributive concerns exhibit a stunningly simple form, simply dividing consumers into two segments: one where the price is the same that would be charged under no segmentation and one with a discount price. In rent markets, we show that consumer-optimal segmentations under sufficiently strong redistributive preferences divide consumers into contiguous segments based on their willingness to pay, having consumers with the same willingness to pay belong to at most two different segments. This allows us to construct a procedure that generates consumer-optimal segmentations under strong redistributive preferences, which is discussed in section 3.4.2.

Related literature. Third-degree price discrimination and its welfare effects are the subject of an extensive literature. Early analysis (Pigou, 1920; Robinson, 1933) and subsequent development (Schmalensee, 1981; Varian, 1985) considered exogenously fixed market segmentations and studied conditions under which such segmentations would increase or decrease total surplus.

This literature has recently undergone a transformation, prompted by both technical innovations in microeconomic theory and the change in character of the practice of price discrimination brought about by the ascent of digital markets. Recent developments incorporate an information design approach to study the welfare impacts of third-degree price discrimination over *all possible* market

segmentations, rather than taking a segmentation as exogenously fixed. Bergemann et al. (2015) analyze a setting with a monopolist selling a single good and characterize attainable pairs of consumer and producer surplus, showing that any distribution of total surplus over consumers and producer that guarantee at least the uniform-price profit for the producer is attainable. In particular, they show that there are typically many consumer-optimal segmentations of a given market. Their analysis has been extended to multi-product settings by Haghpanah and Siegel (2022a,b) and to imperfect competition settings by Elliott, Galeotti, Koh, and Li (2021) and Ali, Lewis, and Vasserman (2022). Hidir and Vellodi (2020) study market segmentation in a setting where the monopolist can offer one from a continuum of goods to each consumer, such that consumers, upon disclosing their information, face a trade-off between being offered their best option and having to pay a fine-tuned price. Finally, Roesler and Szentes (2017) and Ravid, Roesler, and Szentes (2022) study the inverse problem of information design to a buyer who is uncertain about the value of a good. Our paper differs from these by focusing on how surplus is distributed *across* consumers, and by studying consumer-optimal segmentations when different consumers are assigned different welfare weights. We show that, once distributional preferences are taken into account, optimal segmentations might not coincide with consumer-optimal segmentations under uniform welfare weights. When they do, our analysis selects one among the many direct consumer-optimal segmentations established in Bergemann et al. (2015).

Our paper also dialogues with a recent literature on mechanism design and redistribution, most notably with Dworczak et al. (2021) and Akbarpour, Dworczak, and Kominers (2020), who study the design of allocation mechanisms under redistributive concerns; and Pai and Strack (2022), who study the optimal taxation of a good with a negative externality when agents differ on their utility for the good, disutility for the externality and marginal value for money. A key difference in the results obtained in these papers and ours is that, in their settings, redistributive mechanisms are not pareto-efficient: redistribution implies some loss in social surplus. This is not the case in our paper, where optimal redistributive segmentations always maximize total surplus.

Finally, our paper dialogues with Dube and Misra (2022), who study experimentally the welfare implications of personalized pricing implemented through machine learning. The authors find a negative impact of personalized pricing on total consumer surplus, but note that a majority of consumers benefit from price reductions under personalization, pointing that under some inequality-averse weighted welfare functions, data-enabled price personalization might increase

welfare. Their paper shows experimentally how the implementation of market segmentations aimed at maximizing profits might generate, as a by-product, the redistribution of surplus among consumers. Our paper, on the other hand, shows theoretically how consumer-optimal redistributive segmentations might grant additional profits for the firm.

3.2. Model

3.2.1. Setup

A monopolist (he) sells a good to a continuum of mass one of buyers, each of whom can consume at most one unit. We normalize the marginal cost of production of the good to zero. The consumers privately observe their type v, which corresponds to their willingness to pay for the good. We assume that the consumers' type can take a finite number K of possible values $V = \{v_1, \dots, v_K\}$, where $0 < v_1 < \cdots < v_K$. We let $\mathcal{K} := \{1, \dots, K\}$. A *market* μ is a distribution over the valuations. We denote the set of all possible markets:

$$M := \Delta(V) = \left\{\mu \in \mathbb{R}^K \,\middle|\, \sum_{k\in\mathcal{K}} \mu_k = 1 \text{ and } \mu_k \geq 0 \text{ for all } k \in \mathcal{K}\right\}.$$

Price v_k is *optimal for market* $\mu \in M$ if it maximizes the expected revenue of the monopolist when facing market μ, that is:

$$v_k \sum_{i=k}^{K} \mu_i \geq v_j \sum_{i=j}^{K} \mu_i, \quad \forall j \in \mathcal{K}.$$

Let M_k denote the set of markets where price v_k is optimal. It is given by:

$$M_k = \left\{\mu \in M \,\middle|\, v_k \in \operatorname*{arg\,max}_{v_i \in V} v_i \sum_{j=i}^{K} \mu_j\right\},$$

for any $k \in \mathcal{K}$. In the remaining of the paper we will hold an aggregate market fixed and denote it by $\mu^0 \in M$.

Segmentation. The consumers' types are perfectly observed by a social planner (she) who can *segment* consumers, that is, sort consumers into different sub-markets.

The set of possible segmentations of an aggregate market μ^0 is given by:

$$\Sigma(\mu^0) := \left\{\sigma \in \Delta(M) \,\Big|\, \int_{\Delta(M)} \mu\, \sigma(\mathrm{d}\mu) = \mu^0\right\}.$$

Formally, a segmentation is a probability distribution on M which averages to the aggregate market μ^0. The requirement that the different segments generated by a segmentation average to the aggregate market ensures that the segmentation simply sorts existing consumers into different groups, without fundamentally altering the aggregate composition of consumers in a market. This requirement is akin to the Bayes Plausibility condition that is typically used in the Bayesian Persuasion literature (Kamenica and Gentzkow, 2011).

Given a segmentation σ, the monopolist can price differently at each segment μ in the support of σ. A pricing rule is a mapping $p\colon M \to V$. As will become clear in problem 3.4, segments with more than one optimal price play a key role in our results. We focus on the following pricing rule:

$$p(\mu) = \min\left\{\arg\max_{k\in\mathcal{K}}\, v_k \sum_{i=k}^{K} \mu_i\right\}.$$

At each segment, the monopolist charges the smallest price among all optimal prices in that segment. This pricing rule makes the objective of the social planner (stated in equation (P)) upper semi-continuous and ensures the existence of an optimal segmentation[2].

Social objective. The social planner's objective is to maximize a weighted sum of consumers' surplus, with positive weights $\lambda \in \mathbb{R}^K_+$. Each dimension λ_k of the vector λ corresponds to the marginal contribution to social welfare of consumers of type v_k. The surplus of a consumer of type v_k in market μ is given by:

$$U_k(\mu) := \max\left\{0, v_k - p(\mu)\right\}.$$

The weighted consumer surplus on market μ is given by:

$$W(\mu) := \sum_{k\in\mathcal{K}} \lambda_k\, \mu_k\, U_k(\mu),$$

[2]Although technically important, this pricing rule does not impact our results qualitatively. Indeed, any joint distribution of consumers and prices that can be induced by the social planner under this pricing rule could be approximated arbitrarily well by a social planner facing a monopolist who selects among optimal prices in some other way.

for any $\mu \in M$. Hence, for any aggregate market μ^0, the social planner's objective is given by the following maximization program:

$$\max_{\sigma \in \Sigma(\mu^0)} \int_{\Delta(M)} W(\mu)\, \sigma(\mathrm{d}\mu). \tag{P}$$

Given an aggregate market μ^0, a segmentation $\sigma \in \Sigma(\mu^0)$ is *optimal* if it solves (P). We focus on welfare weights that are decreasing on the consumer's willingness to pay, such that $\lambda_k \geq \lambda_{k'}$ for any $k < k' \leq K - 1$, and say that the social planner has *redistributive preferences* if the inequality holds strictly for some $k, k' \in \mathcal{K}$. Under the assumption that consumers with lower willingness to pay are on average poorer than consumers with higher willingness to pay, this amounts to attributing a greater weight to surplus accruing to poorer consumers[3].

Efficiency. Every consumer has a value for the good that is strictly greater than the marginal cost of production. Hence, social surplus is maximized when every consumer buys the good. We say that a market μ is *efficient* if every consumer can buy the good, that is, if the lowest optimal price for the seller at that market allows everyone to consume: $p(\mu) = \min \operatorname{supp}(\mu)$. For a given market μ and Pareto weights λ, the maximum feasible social surplus is thus given by

$$s(\mu) = \sum_{k \in \mathcal{K}} \lambda_k \mu_k v_k.$$

Note that a segmentation of μ achieves $s(\mu)$ if and only if it is efficient. A segmentation σ is *efficient* if it is only supported on efficient markets.

Informational Rents. The profit of the monopolist at market μ is given by:

$$\pi(\mu) = p(\mu) \sum_{k \in C_{p(\mu)}} \mu_k,$$

where $C_p = \{k \in \mathcal{K} \mid v_k \geq p\}$. The profit of the monopolist under segmentation σ is given by:

$$\Pi(\sigma) = \int_{\Delta(M)} \pi(\mu)\, \sigma(\mathrm{d}\mu)$$

Segmenting the aggregate market can only weakly increase the expected profit of the monopolist relative to no segmentation. Therefore, we always have $\Pi(\sigma) \geq \pi(\mu^0)$ for any $\sigma \in \Sigma(\mu^0)$. We say that some segmentation σ grants a *rent* to the

[3] We follow here the approach by Dworczak et al. (2021).

monopolist whenever $\Pi(\sigma) > \pi(\mu^0)$.

Uniformly Weighted Consumer-Optimal Segmentations. If $\lambda_k = \lambda_{k'} > 0$ for all $k, k' \in \mathcal{K}$, program (P) corresponds to the maximization of the total consumer surplus over all possible segmentations. A segmentation that solves this optimization problem is named *uniformly weighted consumer-optimal*. As shown in Bergemann et al. (2015), uniformly weighted consumer-optimal segmentations are (i) efficient—and hence achieve the maximum feasible social surplus—, and (ii) do not grant the monopolist any rent. For an interior aggregate market μ^0, there exists infinitely many uniformly weighted consumer-optimal segmentations. In section 3.4.3, we characterize the set of aggregate markets for which consumer-optimal segmentations *with redistributive preferences* are also uniformly weighted consumer-optimal, thus providing a natural way to select among these segmentations for such markets.

3.2.2. Discussion of the model

Information provision as segmentation. In digital markets, information provision about consumers often occurs through the assignment of *labels* to different consumers. Indeed, one could think of a model in which the social planner adopts a signal structure $\ell\colon V \to \Delta(L)$, where L is a set of labels. The meaning of each label is then pinned down by the social planner's strategy, and the monopolist optimally chooses different prices for consumers with different labels.

Such a model is equivalent to ours. Indeed, any segmentation $\sigma \in \Sigma(\mu^0)$ can be implemented by some signal structure ℓ, and any signal structure ℓ implements some segmentation $\sigma \in \Sigma(\mu^0)$. The approach of working directly in the space of feasible distributions over markets rather than in the space of labeling strategies is standard in the information design literature (Kamenica and Gentzkow, 2011).

Continuum of consumers. While we consider a setting with a continuum of consumers, our model is equivalent to one in which there is a discrete number of consumers, with types independently distributed according to μ^0. Under this interpretation, the social planner commits ex-ante to an information structure σ to inform the monopolist, which defines the distribution of posterior beliefs μ that the monopolist will form upon facing each consumer.

3.3. Three values case

In this section, we illustrate our model and some of the results from the following sections in the simple three values case.

Setup. Let's consider three types, $v_1 = 1$, $v_2 = 2$ and $v_3 = 3$. We can conveniently depict the set of markets M as the two-dimensional unit simplex (see Mas-Colell, Whinston, and Green, 1995, p.169). It is depicted in figure 3.1, where each vertex of the simplex represents a degenerate market on a value $v \in V$, denoted by the Dirac measure δ_v.

In the left panel of figure 3.1 are drawn the three different price regions M_1, M_2 and M_3. The points in each of the regions correspond to the markets for which each of the different prices $\{1, 2, 3\}$ are optimal for the monopolist[4]. The border between two adjacent regions represents markets for which there are more than one optimal price. Given pricing rule p, the price charged in such markets is the lowest amongst the optimal.

In the right panel, an aggregate market $\mu^0 = (0.3, 0.4, 0.3)$ is represented, which is in the interior of the region M_2, meaning that v_2 is a strictly optimal price for μ^0. Two possible segmentations are depicted: the one in green dashed lines, that segments μ^0 into the three degenerate markets (thus implementing first-degree price discrimination); and the one in black dotted lines, that segments μ^0 into three segments: μ', containing types all three types and being priced v_1; μ'', containing only types v_2 and v_3 and being priced v_2; and μ''', containing all three types and being priced v_3.

Any splitting of μ^0 into a set of points $S \subset M$ represents a feasible segmentation, as long as $\mu^0 \in \text{co}(S)$[5]. A segmentation is optimal given weights $(\lambda_1, \lambda_2, \lambda_3)$, with $\lambda_1 \geq \lambda_2 \geq \lambda_3$, if it maximizes the sum of weighted consumer surplus over all segments generated. Note that consumers of type v_1 never get any consumer surplus (since the monopolist never charges a price lower than their willingness to pay), such that the optimal segmentation trades-off surplus obtained by types v_2 and v_3. We will focus, without loss of generality, on direct segmentations, i.e. segmentations in which there is not more than one segment with a given price.

General properties of optimal segmentations. A first step for finding the optimal segmentation of μ^0 is to observe that any optimal segmentation must be

[4]Formally, for any k, $M_k = \text{cl}(p^{-1}(v_k))$, where $\text{cl}(S)$ denotes the topological closure of a generic set S.

[5]For any set S, $\text{co}(S)$ denotes the convex hull of S

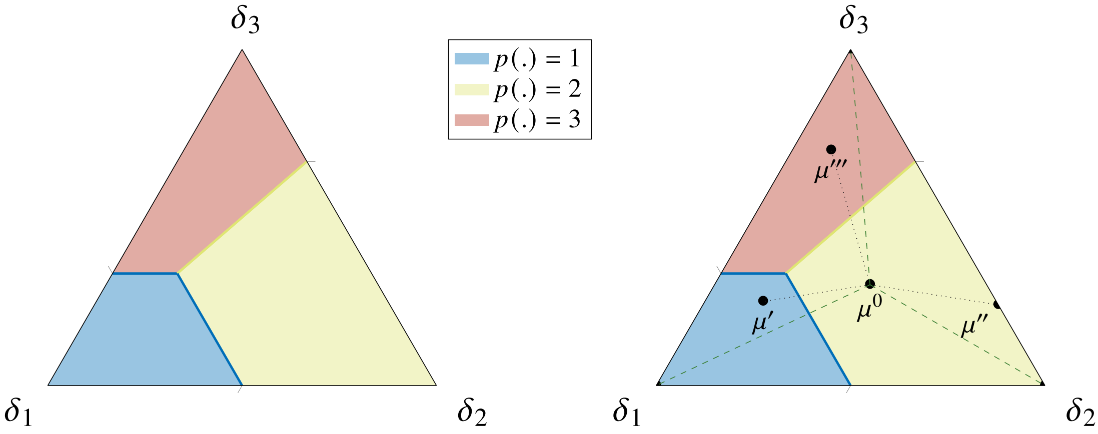


Figure 3.1: The Simplex representing M and two feasible segmentations.

efficient. To see that, consider the black dotted segmentation in the right panel of figure 3.1. Both μ' and μ'' are efficient, since all the consumers in these segments are able to buy the good. The remaining segment μ''', however, is not efficient, as it contains some consumers with type v_1 and v_2 who are not able to consume under that segment's price. One could solve that by re-segmenting μ''' in the following way: creating a segment μ_b''' containing all of the types v_1 and v_2 and some of the types v_3 that used to belong to μ''', and another segment δ_3 containing only the remaining types v_3. Note that the amount of type v_3 in μ_b''' can be adjusted to ensure that this segment will have price v_1. That way, both of the resulting segments will be efficient. Furthermore, this re-segmentation of μ''' *unambiguously* increases consumer welfare, since it has no impact on the welfare of consumers in μ' and μ'' and (weakly) increases the surplus of every consumer previously belonging to μ'''.

Indeed, a welfare-increasing segmentation can be performed to any inefficient market. This narrows down the search for an optimal segmentation, as we know that it must be supported *only* on efficient segments. The left panel of figure 3.2 depicts, in orange, the efficient markets. These are: the degenerate market δ_3; the set of markets in region M_2 that have no consumer with value 1; and the entire region M_1.

We can further note that, in an optimal segmentation, the segment with price v_1 must not belong to the interior of region M_1. To see that, consider the right panel of figure 3.2. In it are depicted two segmentations: σ_a, which splits μ^0 into μ_a and μ', and σ_b, which splits μ^0 into μ_b and μ'. Segmentation σ_b is always preferred over σ_a for two reasons. First, μ_b has a higher share of types v_2 and v_3 than μ_a. Since these are the only two types that are extracting surplus on the segment whose price is v_1, having a higher share of them increases the social planner's objective. Second, μ_b is "closer" to μ^0, which means that $\sigma_b(\mu_b) > \sigma_a(\mu_a)$. That means

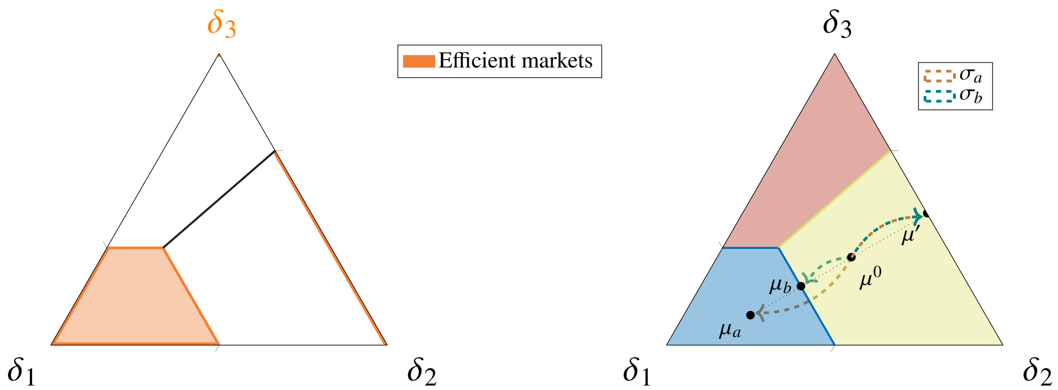


Figure 3.2: Efficient markets and segmentations.

that segmentation σ_b is able to include a bigger mass of consumers in the segment where they will extract the largest surplus, thus also increasing the social planner's objective.

The argument outlined above illustrates how every segmentation generating a segment on the interior of region M_1 must be dominated by some segmentation that instead generates a segment on the boundary of regions M_1 and M_2. This amounts to saying that any optimal segmentation must include a segment in which the monopolist is indifferent between charging price v_1 or charging some other price. The intuition for that is simple: if the monopolist strictly prefers to charge price v_1 in that segment, then there's still room for "fitting" other types in that segment in a Pareto improving way.

Uniformly weighted consumer-optimal segmentations. We begin by considering the case where $\lambda_1 = \lambda_2 = \lambda_3$. The left panel of figure 3.3 depicts three different segmentations, σ_a, σ_b and σ_c, each of them generating one segment with price v_1 and one segment with price v_2. All of these three segmentations are uniformly weighted consumer-optimal. This follows from the fact that i) they maximize total (consumer + producer) surplus, since they are all efficient, and ii) the monopolist does not get any of the surplus that is created from the segmentation [6].

Indeed, there are uncountably many uniformly weighted consumer-optimal segmentations of μ^0. All of these are equivalent in that they maximize total

[6] One way of seeing this is as follows: A decision-maker strictly benefits from observing a piece of information if, as a result of this observation, she is able to make better decisions than she would have made absent this information. In our setting, this amounts to the monopolist being able to, as a result of the segmentation, choose *different* prices than the uniform price, at markets in which these different prices are *strictly* preferred over the uniform price. Since price v_2 belongs to the set of optimal prices in every segment generated by the segmentations in figure 3.3, the monopolist does not strictly benefit from them.

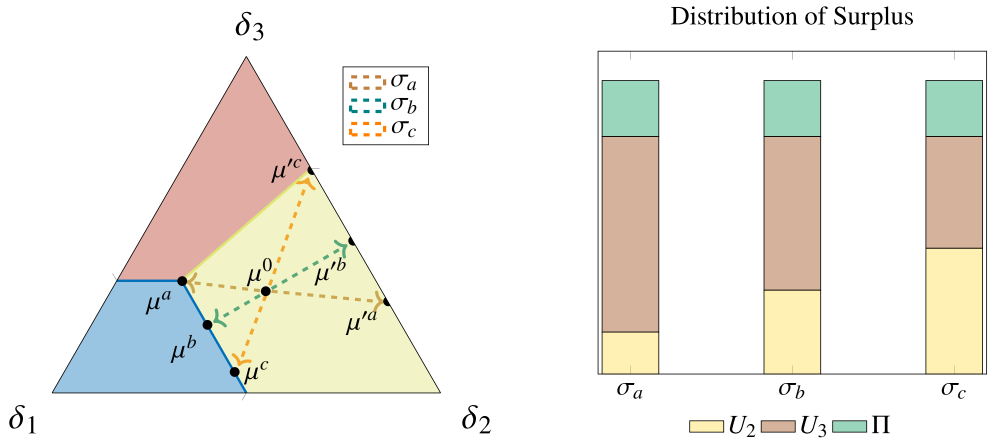


Figure 3.3: Uniformly Weighted Consumer-Optimal Segmentations.

consumer surplus, but they are not equivalent in how they distribute such surplus *across* consumers. This can be seen in the right panel of figure 3.3: while the three segmentations of the left panel induce the same profit for the monopolist and the same total consumer surplus, σ_c induces greater surplus for consumers of type v_2 than the other segmentations. This is so because, among the segments priced at v_1, μ_c is the one that includes the most consumers of type v_2, who can then benefit from a low price.

Consumer-Optimal segmentations under redistributive preferences. Let's now consider the case when $\lambda_2 > \lambda_3$. Among the segmentations depicted in the left panel of figure 3.3, segmentation σ_c is now preferred over σ_a and σ_b. But is it optimal? One way of increasing the surplus of consumers of type v_2 further is to exchange consumers between the two segments generated by σ_c: by exchanging the remaining consumers of type v_3 that are present in μ^c against some of the consumers of type v_2 present in μ'^c, one can increase the amount of types v_2 that pay a low price. While this exchange increases the surplus of types v_2, it dramatically decreases the surplus of types v_3, since now there are sufficiently many of them in segment μ'^c for the monopolist to want to increase the price charged at that segment. This would lead to a segmentation that is no longer uniformly weighted consumer-optimal: the price increase in segment μ'^c would cause some of the surplus that was previously captured by consumers of type v_3 to now be granted to the monopolist instead. The result below establishes when this exchange is desirable from the social planner's perspective.

Result 1. *Let* $\mu^0 = (0.3, 0.4, 0.3)$. *Then, the two following assertions are satisfied:*

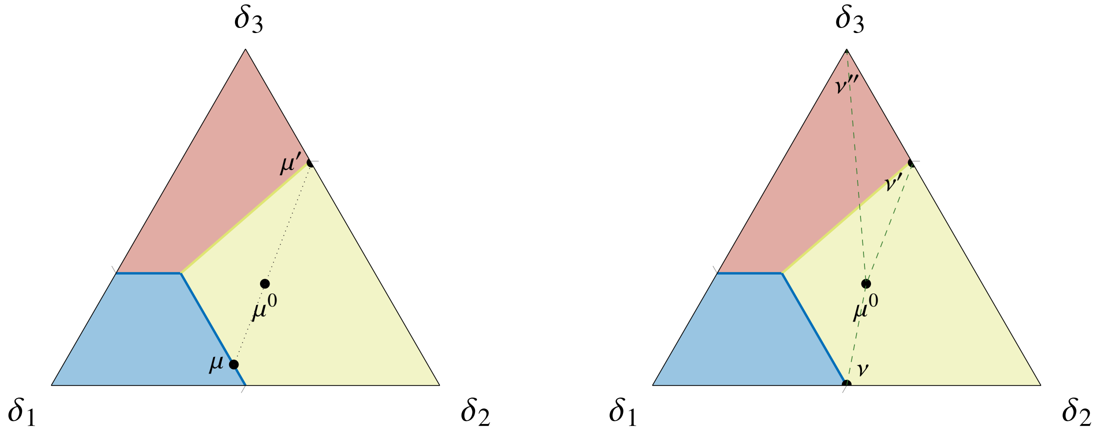


Figure 3.4: Optimal Segmentations with Redistributive Preferences.

(i) If the inequality

$$\frac{\lambda_2}{\lambda_3} < \frac{v_3 + v_2 - v_1}{v_2 - v_1},$$

is satisfied, then the consumer-optimal segmentation under redistributive preferences is also uniformly weighted consumer-optimal and generates two segments. One supported on $\{v_1, v_2, v_3\}$ *and the other one supported on* $\{v_2, v_3\}$*. This segmentation is represented in the left panel of figure 3.4;*

(ii) If the inequality

$$\frac{\lambda_2}{\lambda_3} > \frac{v_3 + v_2 - v_1}{v_2 - v_1},$$

is satisfied, then the consumer-optimal segmentation under redistributive preferences is not *uniformly weighted consumer-optimal and generates three segments. The first one is supported on* $\{v_1, v_2\}$*, the second is supported on* $\{v_2, v_3\}$*, and the third is supported on* $\{v_3\}$*. This segmentation is represented in the right panel of figure 3.4.*

An important consequence of this result is that if the social planner's preferences are sufficiently redistributive, meaning that λ_2 is sufficiently greater than λ_3, the optimal segmentation might give a *rent* (i.e. an additional profit) to the monopolist. By packing more consumers with lower types together, the social planner also makes higher types more distinguishable, thus allowing the monopolist to raise their prices. The above example illustrates the main argument of the paper: while market segmentation can redistribute surplus without any loss of efficiency, sometimes raising the surplus of poorer consumers can only be done if some of the surplus from richer consumers is granted to the monopolist.

However, not every aggregate market requires the granting of rents to the

monopolist in order to satisfy redistributive objectives. Consider for instance the aggregate market $\mu^0 = (0.2, 0.65, 0.15)$, represented in the left panel of figure 3.5. The optimal segmentation of this market given *any* preferences $\lambda_2 \geq \lambda_3$ is the one depicted in the figure: it always generates a segment with $\{v_1, v_2\}$ and another one with $\{v_2, v_3\}$, and this segmentation is always uniformly weighted consumer-optimal. On this aggregate market, satisfying a redistributive objective never requires granting rents to the monopolist because it contains sufficiently many consumers of type v_2, such that even after pooling as many as possible of them with types v_1 in segment μ, there are still sufficiently many types v_2 left to ensure that types v_3 will not be over-represented in segment μ'.

The result below characterizes the set of aggregate markets that, under a sufficiently strong redistributive motive, would require granting rents to the monopolist. We denote this set as the *rent region*.

Result 2. *The rent region is give by*

$$\operatorname{int}\Big(\operatorname{co}\Big(\big\{\delta_3, \mu^{123}, \mu^{12}, \mu^{23}\big\}\Big)\Big).$$

This result is illustrated in the right panel of figure 3.5, where the rent region is depicted in orange. Equivalently, the complement of this set denotes the aggregate markets for which any redistributive objective can be met without granting rents to the monopolist — that is, while maximizing total consumer surplus—. We call this set the *no-rent region*. The following section generalizes the insights

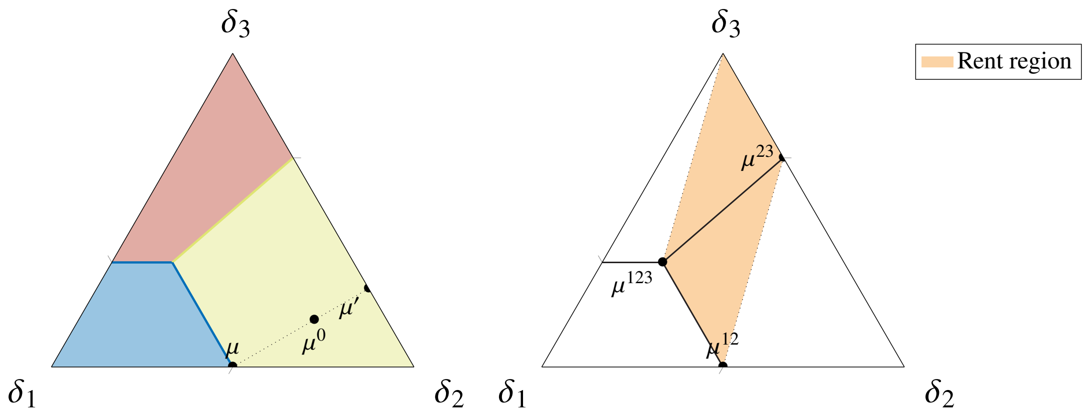


Figure 3.5: Rent Region.

presented through this example. Section 3.4.1 generalizes the fact that optimal segmentations are efficient and include discount segments supported at markets at which the monopolist is indifferent between more than one price, while section 3.4.2 establishes properties of optimal segmentations when the redistributive motive

is sufficiently strong and shows how to construct optimal segmentations in this case. Finally, section 3.4.3 characterizes generally the no-rent and rent regions and shows that optimal segmentations for markets belonging to the no-rent region exhibit a very simple form, with only one discount segment and one uniform price segment.

3.4. Optimal segmentations

We now turn to the analysis of the general case. In section 3.4.1 we derive general properties of optimal segmentations — that is, characteristics that are present in optimal segmentations given any decreasing welfare weights λ. Section 3.4.2 then constructs optimal segmentations under strongly redistributive preferences: when the weight assigned to lower types is sufficiently larger than the weight assigned to higher types. Finally, we characterizes the set of aggregate markets for which satisfying a redistributive objective might require granting additional profits to the monopolist in section 3.4.3.

3.4.1. General properties

Efficient segmentations. Our first result echoes our analysis of efficiency in the three-value case and establishes that i) we can always restrict ourselves to efficient segmentations—as long as the weights are non-negative; ii) if the weights are all strictly positive (i.e. if $\lambda_K > 0$ under our assumption of decreasing weights), only efficient segmentations can be optimal.

Proposition 12. *For any aggregate market μ^0 and any weights $\lambda \in \mathbb{R}_+^K$ (not necessarily decreasing), there exists an efficient optimal segmentation of μ^0. Furthermore, if every weight is strictly positive, then any optimal segmentation is efficient.*

Proof. This result is a direct consequence of Proposition 1 in Haghpanah and Siegel (2022b)—which itself follows from the proof of Theorem 1 in Bergemann et al. (2015). □

This result relies on the fact that any inefficient market can be segmented in a Pareto improving manner, that is, in a way that weakly increases the surplus of all consumers. Hence, as long as the social planner does not assign a negative weight to any consumer, there must be an efficient optimal segmentation. Proposition 12 thus implies that segmenting in a redistributive manner never comes at the expense of efficiency.

Direct segmentations. A segmentation σ is *direct* if all segments in σ have different prices, that is, if for any $\mu, \mu' \in \text{supp}(\sigma)$, $p(\mu) \neq p(\mu')$. Our next lemma shows that it is without loss of generality to focus on direct segmentations.

Lemma 12. *For any aggregate market μ^0 and any segmentation $\sigma \in \Sigma(\mu^0)$, there exists a direct segmentation $\sigma' \in \Sigma(\mu^0)$ such that,*

$$\int_{\Delta(M)} W(\mu)\,\sigma(\mathrm{d}\mu) = \int_{\Delta(M)} W(\mu)\,\sigma'(\mathrm{d}\mu).$$

Proof. See appendix C.1.1 □

We further show that there always exists an optimal and direct segmentation that is only supported on the boundaries of price regions $\{M_k\}_{k\in\mathcal{K}}$. Let $\mathcal{K}^0 := \{k \in \mathcal{K} \mid v_k \in \text{supp}(\mu^0)\}$ be the set of indices of consumers' types supported by μ^0.

Lemma 13. *For any aggregate market μ^0 that is not efficient, there exists an optimal direct segmentation supported on boundaries of sets $\{M_k\}_{k\in\mathcal{K}^0}$.*

Proof. See appendix C.1.2. □

This result implies that we can restrict without loss of generality to finitely supported segmentations.

3.4.2. Strongly redistributive social preferences

In this section, we derive some characteristics of the optimal segmentation when the social planner's preferences are *strongly redistributive*, that is, when the weights λ are strongly decreasing on the type v.

Definition 7. *The weights λ are κ-*strongly redistributive *if, for any $k < k' \leq K-1$, $\frac{\lambda_k}{\lambda_{k'}} \geq \kappa$.*

That is, a social planner exhibits κ-strongly redistributive preferences (κ-SRP) if the weight she assigns to a consumer of type v_k is at least κ times larger than the weight she assigns to any consumer of type greater than v_k.

Let us define the *dominance* ordering between any two sets.

Definition 8. *Let $X, Y \subset \mathbb{R}$. The set X* dominates *Y, denoted $X \geqslant_D Y$, if for any $x \in X$ and any $y \in Y$, $x \geq y$.*[7]

We can now state the main result of this section.

[7]Note that this definition of dominance is stronger than the strong set order in Topkis (1998).

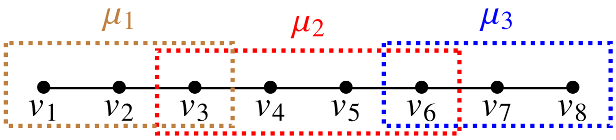


Figure 3.6: Structure of optimal segmentations under strong redistributive preferences.

Proposition 13. *For any aggregate market μ^0 in the interior of M, there exists $\underline{\kappa}$ such that if λ's are $\underline{\kappa}$-strongly redistributive, then for any optimal direct segmentation $\sigma \in \Sigma(\mu^0)$ and any markets $\mu, \mu' \in$ supp(σ), $\mu \neq \mu'$: either* $\text{supp}(\mu) \geqslant_D \text{supp}(\mu')$ *or* $\text{supp}(\mu') \geqslant_D \text{supp}(\mu)$.

Proof. See appendix C.2.1 □

The result stated above establishes that, when the social planner's preferences exhibit a sufficiently strong taste for redistribution, optimal segmentations divide the type space V into contiguous overlapping intervals, with the overlap between any two segments being composed of at most one type. The following corollary is a direct consequence of proposition 13:

Corollary 5. *For any aggregate market μ^0 in the interior of M, there exists $\underline{\kappa}$ such that if λ's are $\underline{\kappa}$-strongly redistributive, then for any optimal direct segmentation $\sigma \in \Sigma(\mu^0)$, any market $\mu \in$ supp(σ) and any k such that* $\min\{\text{supp}(\mu)\} < v_k < \max\{\text{supp}(\mu)\}$*: $\sigma(\mu)\mu_k = \mu^0_k$.*

The above result states that any segment μ belonging to a segmentation that is optimal under strong redistributive preferences contains *all* of the consumers with types strictly in-between $\min\{\text{supp}(\mu)\}$ and $\max\{\text{supp}(\mu)\}$. Together with proposition 13, it implies that, under κ-SRP optimal segmentations, every consumer type v will belong to *at most* two segments: either it will belong to the interior of the support of a segment μ, such that all consumers of this type have surplus $v - \min(\text{supp}(\mu))$, or it will be the boundary type between two segments μ and μ', such that a fraction of these consumers (those belonging to segment μ) gets surplus $v - \min(\text{supp}(\mu))$ and the rest gets no surplus. The structure of optimal segmentations under strong redistributive preferences is illustrated in figure 3.6.

These results, along with proposition 12, completely pin down the κ-SRP optimal direct segmentation. One can construct it by employing the following procedure, presented as follows through steps:

- *Step i)* Start by creating a segment — call it μ_a — with all consumers of type v_1.

- *Step ii)* Proceed to including in μ_a, successively, all consumers of type v_2, then all of the types v_3, and so on. From proposition 12 we know that μ_a must be efficient, meaning that we must have $p(\mu_a) = v_1$. As such, the process of inclusion of types higher than v_1 must be halted at the point in which adding a new consumer in μ_a would result in v_1 no longer being an optimal price in this segment. We denote as $v_{(a|b)}$ the type that was being included when the process was halted.
- *Step iii)* Create a new segment — call it μ_b — with all of the remaining types $v_{(a|b)}$.
- *Step iv)* Proceed to including in μ_b, successively, all of consumers of type $v_{(a|b)+1}$, then all of the types $v_{(a|b)+2}$, and so on. Halt this process at the point in which adding a new consumer in μ_b would result in $v_{(a|b)}$ no longer being an optimal price in this segment. We denote as $v_{(b|c)}$ the type that was being included when the process was halted.
- *Step v)* Create a new segment with all of the remaining types $v_{(b|c)}$. Repeat the process described in the last steps until every consumer has been allocated to a segment.

3.4.3. Optimal segmentations and informational rents

This section explores the question of when does an optimal segmentation maximize total consumer surplus or, conversely, when it grants a rent for the monopolist.

Say that an aggregate market μ^0 belongs to the *rent region* if there exists some $\underline{\kappa}$ such that if the social planner has $\underline{\kappa}$-strongly redistributive preferences, the optimal segmentation grants a rent to the monopolist. Conversely, denote *no-rent region* the set of aggregate markets for which any optimal segmentation with redistributive preferences also maximizes total consumer surplus.

Before we characterize the rent and no-rent regions, we define a particular segmentation, which we will call σ^{NR}:

Definition 9. *Let μ^0 be an aggregate market with uniform price v_u. Call σ^{NR} the segmentation that splits μ^0 into two segments μ^s and μ^r, such that:*

$$\mu^s = \left(\frac{\mu_1^0}{\sigma}, \frac{\mu_2^0}{\sigma}, \dots, \mu_u^s, 0, \dots, 0\right),$$

$$\mu^r = \left(0, 0, \dots, \mu_u^r, \frac{\mu_{u+1}^0}{1-\sigma}, \dots, \frac{\mu_K^0}{1-\sigma}\right),$$

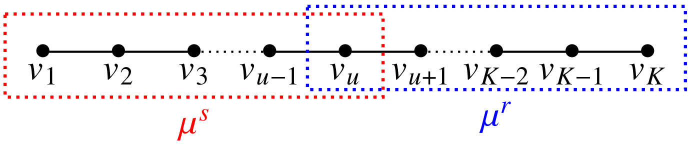


Figure 3.7: Segmentation σ^{NR}.

where $\mu_u^s = v_1/v_u$, $\mu_u^r = (\mu_u^0 - \sigma\mu_u^s)/(1-\sigma)$ *and* $\sigma = (v_u \sum_{i=1}^{u-1} \mu_i^\star)/(v_u - v_1)$.

Segmentation σ^{NR} is very simple and generates only two segments: one pooling all the consumers who would not buy the good on the unsegmented market (those with type lower than v_u) and another one pooling all the consumers who would buy the good on the unsegmented market (those with type higher than v_u). Under segmentation σ^{NR}, the only consumer type that gets assigned to two different segments is v_u.

Proposition 14. *An aggregate market μ^0 belongs to the no-rent region if and only if σ^{NR} is an efficient segmentation of μ^0.*

Proof. See appendix C.3.1. □

Proposition 14 establishes a simple criterion that defines whether an aggregate market belongs to the no-rent region: it suffices to check if, under σ^{NR}, $p(\mu^s) = v_1$ and $p(\mu^r) = v_u$. Whenever this is not true, the aggregate market belongs to the rent region.

Corollary 6. *Consider an aggregate market μ^0. If σ^{NR} is not an efficient segmentation of μ^0, then there exists $\underline{\kappa}$ such that, if welfare weights λ are $\underline{\kappa}$-strongly redistributive, any optimal segmentation grants a rent to the monopolist.*

The intuition for the results above is as follows. A market belongs to the no-rent region if, given any redistributive preferences, its optimal segmentation maximizes total consumer surplus. On one hand, we know from proposition 13 that, under strong redistributive preferences, optimal segmentations divide the type space into overlapping intervals, with the overlap between two segments being comprised of at most one type. On the other hand, we have as a necessary and sufficient condition for total consumer surplus to be maximized that the segmentation is i) efficient and ii) the uniform price v_u is an optimal price at *every* segment generated by this segmentation. Condition i) ensures that total surplus is maximized, while condition ii) ensures that producer surplus is kept at it's uniform price level, meaning that all of the surplus created by the segmentation goes to consumers. Since condition ii) can only be satisfied if type v_u belongs in the support of all segments, we get

that the conditions for optimality under strong redistributive preferences and for total consumer surplus to be maximized can only be simultaneously met by a segmentation that only generates two segments, with the overlap in the support of both segments being comprised of v_u.

Such a segmentation indeed maximizes total consumer surplus if it is efficient and if v_u is an optimal price in both segments. This is the case if v_1 and v_u are both optimal optimal prices on the lower segment, and if v_u is an optimal price in the upper segment. Segmentation σ^{NR} is the *only* segmentation that can potentially satisfy all of these conditions at once, as it includes in the lower segment the exact proportion of types v_u that would make the monopolist indifferent between charging a price of v_1 or v_u. As such, segmentation σ^{NR} maximizes total consumer surplus if and only if it is efficient.

Corollary 7. *If an aggregate market μ^0 belongs to the no-rent region, then σ^{NR} is its only direct consumer-optimal segmentation under any redistributive preferences.*

This result establishes that, for markets in the no-rent region, optimal segmentations have an extremely simple structure: they only generate a discount segment with price v_1, pooling all the types who would not consume under the uniform price and some of the types v_u, and a residual segment with price v_u, containing all of the remaining consumers. Furthermore, this segmentation must be optimal under *any* decreasing welfare weights λ. As such, this result selects for the markets belonging to the no-rent region one among the many uniformly weighted consumer-optimal segmentations that were outlined in Bergemann et al. (2015).

Due to the structure of segmentation σ^{NR}, all of the surplus that is generated by the segmentation is given to consumers with types below or equal to v_u, all of which get the maximum surplus they could potentially get. Since it is impossible to raise the surplus of any type below v_u, and impossible to raise the surplus of types above v_u without redistributing from lower to higher types, this segmentation must be optimal whenever the weights assigned to different consumers are (weakly) decreasing on the type.

The results in this section establish that there are essentially two types of markets: those for which redistribution can be done only within consumers, while keeping total consumer surplus maximal, and those for which increasing the surplus of lower types past a certain point necessarily decreases the total pie of surplus accruing to consumers and grants additional profits to the monopolist.

Bibliography

Akbarpour, M., P. Dworczak, and S. D. Kominers (2020): “Redistributive Allocation Mechanisms,” *Working Paper*.

Ali, S. N., G. Lewis, and S. Vasserman (2022): “Voluntary Disclosure and Personalized Pricing,” *The Review of Economic Studies*, rdac033.

Bergemann, D., B. Brooks, and S. Morris (2015): “The Limits of Price Discrimination,” *American Economic Review*, 105, 921–57.

Dube, J.-P. and S. Misra (2022): “Personalized Pricing and Consumer Welfare,” *Journal of Political Economy*, Forthcoming.

Dworczak, P., S. D. Kominers, and M. Akbarpour (2021): “Redistribution Through Markets,” *Econometrica*, 89, 1665–1698.

Elliott, M., A. Galeotti, A. Koh, and W. Li (2021): “Market Segmentation through Information,” *Working Paper*.

Haghpanah, N. and R. Siegel (2022a): “The Limits of Multi-Product Price Discrimination,” *American Economic Review: Insights*, Forthcoming.

——— (2022b): “Pareto Improving Segmentation of Multi-Product Markets,” *Journal of Political Economy*, Forthcoming.

Hidir, S. and N. Vellodi (2020): “Privacy, Personalization, and Price Discrimination,” *Journal of the European Economic Association*, 19, 1342–1363.

Kamenica, E. and M. Gentzkow (2011): “Bayesian Persuasion,” *American Economic Review*, 101, 2590–2615.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995): *Microeconomic Theory*, New York: Oxford University Press.

Pai, M. and P. Strack (2022): “Taxing Externalities Without Hurting the Poor,” *Working Paper*.

Pigou, A. C. (1920): *The Economics of Welfare*, London: Macmillan.

Ravid, D., A.-K. Roesler, and B. Szentes (2022): “Learning before Trading: On the Inefficiency of Ignoring Free Information,” *Journal of Political Economy*, 130, 346–387.

Robinson, J. (1933): *The Economics of Imperfect Competition*, London: Macmillan.

Roesler, A.-K. and B. Szentes (2017): "Buyer-Optimal Learning and Monopoly Pricing," *American Economic Review*, 107, 2072–80.

Schmalensee, R. (1981): "Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination," *American Economic Review*, 71, 242–247.

Topkis, D. M. (1998): *Supermodularity and Complementarity*, Princeton University Press.

Varian, H. R. (1985): "Price Discrimination and Social Welfare," *American Economic Review*, 75, 870–875.

Appendices

A. Mathematical appendix for chapter 1

A.1. Proof of proposition 1

Let (σ, τ) be an arbitrary t-pass-fail mechanism. First, remark that it would never be optimal for the designer to choose a cutoff $t < 0$ by lemma 1. Thus, we can optimize over the class of pass-fail mechanisms with positive cutoffs without loss of optimality. The payoff of the designer under any t-pass-fail with $t \geq 0$ is given by the following function of the single variable t:

$$V(t) = \begin{cases} t\,(F(t) - F(\theta(t))) + \displaystyle\int_t^{\bar{\theta}} \theta f(\theta)\,\mathrm{d}\theta & \text{if } t \in [0, \bar{\theta}[\\ t\,(1 - F(\theta(t))) & \text{if } t \in [\bar{\theta}, \bar{\theta} + \sqrt{2/\gamma}] \end{cases}.$$

Remembering that $\theta(t) = t - \sqrt{2/\gamma}$ we can deduce that the function $V\colon [0, \bar{\theta} + \sqrt{2/\gamma}] \to \mathbb{R}$ is continuously differentiable with derivative given by:

$$V'(t) = \begin{cases} F(t) - F(\theta(t)) - t f(\theta(t)) & \text{if } t \in [0, \bar{\theta}[\\ 1 - F(\theta(t)) - t f(\theta(t)) & \text{if } t \in [\bar{\theta}, \bar{\theta} + \sqrt{2/\gamma}] \end{cases}.$$

If an optimal threshold t^*_γ exists, it must therefore satisfy the following first-order condition:

$$V'(t) = 0. \tag{FOC}$$

Consider now the following function:

$$\psi(t) = \begin{cases} t - \dfrac{F(t) - F(\theta(t))}{f(\theta(t))} & \text{if } t \in [0, \bar{\theta}[\\[2ex] t - \dfrac{1 - F(\theta(t))}{f(\theta(t))} & \text{if } t \in [\bar{\theta}, \bar{\theta} + \sqrt{2/\gamma}] \end{cases}.$$

It is easy to see that the equation (FOC) admits the same solution as the equation

$$\psi(t) = 0. \tag{FOC'}$$

if it exists. The function ψ is continuous over the interval $[0, \bar{\theta} + \sqrt{2/\gamma}]$ and satisfies

$$\psi(0) = -\frac{F(0) - F\Big(-\sqrt{\frac{2}{\gamma}}\Big)}{f\Big(-\sqrt{\frac{2}{\gamma}}\Big)} < 0$$

as well as

$$\psi\big(\bar{\theta} + \sqrt{2/\gamma}\big) = \bar{\theta} + \sqrt{\frac{2}{\gamma}} > 0$$

Hence, by the Intermediate Value Theorem, a solution to equation (FOC') must exist in the interval $]0, \bar{\theta} + \sqrt{2/\gamma}[$. Remark also that, under assumption 1, the function ψ is strictly increasing on that interval (whenever f is not the uniform distribution), since:

$$\psi'(t) = \left(1 - \underbrace{\frac{f(t)}{f(\theta(t))}}_{\leq 1}\right) + 1 - \left(\underbrace{\frac{f'(\theta(t))}{f(\theta(t))}}_{\leq 0}\right)\left(\underbrace{\frac{F(t) - F(\theta(t))}{f(\theta(t))}}_{>0}\right) \geq 0,$$

for all $t \in [0, \bar{\theta}]$ and

$$\psi'(t) = \left(\underbrace{\frac{f'(\theta(t))}{f(\theta(t))}}_{\leq 0}\right)\left(\underbrace{\frac{1 - F(\theta(t))}{f(\theta(t))}}_{>0}\right) > 0,$$

for all $t \in [\bar{\theta}, \bar{\theta} + \sqrt{2/\gamma}]$. This implies that the solution to equation (FOC') must be unique in the interval $]0, \bar{\theta} + \sqrt{2/\gamma}[$.

Let us prove additionally that ψ must achieve its unique zero in the interval $]0, \sqrt{2/\gamma}[$ for any $\gamma > 0$. Remark first that:

$$\psi(\sqrt{2/\gamma}) = \begin{cases} \sqrt{\frac{2}{\gamma}} - \frac{1 - F(0)}{f(0)} & \text{if } \gamma \in]0, 1/c(\bar{\theta}, 0)[\\ \sqrt{\frac{2}{\gamma}} - \frac{F\left(\sqrt{\frac{2}{\gamma}}\right) - F(0)}{f(0)} & \text{if } \gamma \in [1/c(\bar{\theta}, 0), +\infty[\end{cases}$$

Under assumption 1 we must have:

$$\frac{\mathrm{d}}{\mathrm{d}\gamma}\psi(\sqrt{2/\gamma}) = \begin{cases} \overbrace{-\dfrac{1}{\gamma\sqrt{2\gamma}}}^{<0} & \text{if } \gamma \in \,]0, 1/c(\bar{\theta}, 0)[\\ -\dfrac{1}{\gamma\sqrt{2\gamma}}\left(1 - \underbrace{\dfrac{f\left(\sqrt{\frac{2}{\gamma}}\right)}{f(0)}}_{\leq 1}\right) & \text{if } \gamma \in [1/c(\bar{\theta}, 0), +\infty[\end{cases}$$

which implies that $\psi(\sqrt{2/\gamma})$ is strictly decreasing in γ. Moreover, we have:

$$\lim_{\gamma \to 0^+} \psi(\sqrt{2/\gamma}) = +\infty$$

as well as

$$\lim_{\gamma \to +\infty} \psi(\sqrt{2/\gamma}) = 0$$

All those facts put together imply that $\psi(\sqrt{2/\gamma}) > 0$ for all $\gamma > 0$. Whence, again by the Intermediate Value Theorem, the optimal cutoff t^*_γ must lie in the interval $]0, \sqrt{2/\gamma}[$ for any $\gamma > 0$. This also implies that $\theta^*_\gamma = \theta(t^*_\gamma) \in \,]-\sqrt{2/\gamma}, 0[$ for any γ.

A.2. Additional proofs for theorem 1

A.2.1. Proof of lemma 1

We prove successively each of the properties:

(i) Let us first prove that if σ is optimal it must assign zero probability to strictly negative types. Assume that σ is optimal and $\sigma(t) > 0$ for some $t \in [\underline{t}, 0[$. Let $\tau_\sigma(\theta) \in \arg\max_{t \in T} \sigma(t) - \gamma c(t, \theta)$ and define the sets

$$\Theta^-_\sigma = \left\{\theta \in \Theta \,|\, \tau_\sigma(\theta) < 0\right\},$$

and

$$\Theta^+_\sigma = \left\{\theta \in \Theta \,|\, \tau_\sigma(\theta) \geq 0\right\}.$$

Remark that we can always write designer's payoff as

$$V(\sigma) = \int_{\Theta_\sigma^-} \tau_\sigma(\theta)\, \sigma(\tau_\sigma(\theta)) f(\theta)\, \mathrm{d}\theta + \int_{\Theta_\sigma^+} \tau_\sigma(\theta)\, \sigma(\tau_\sigma(\theta)) f(\theta)\, \mathrm{d}\theta.$$

By definition, the first integral is negative while the second is positive. Now, define $\varsigma(t)$ to be such that $\varsigma(t) = 0$ for any $t \in]-\infty, 0[$ and $\varsigma(t) = \sigma(t)$ for any $t \in [0, +\infty[$. If Θ_σ^- is empty, then moving from σ to ς would not change designer's expected payoff and would still be optimal. If Θ_σ^- is non-empty, then moving from σ to ς brings the value of the first integral to zero and could have only two effects on the second integral since types in Θ_σ^+ have unchanged incentives: First, either no type in Θ_σ^- would invest in a positive final type under ς implying $\Theta_\varsigma^+ = \Theta_\sigma^+$ and thus keeping the value of the second integral unchanged while bringing the value of the first integral to zero. Second, some type in Θ_σ^- could be willing to invest in a strictly positive type under ς implying $\Theta_\sigma^+ \subset \Theta_\varsigma^+$ and thus increasing the value of the second integral. In both those cases, we have $V(\varsigma) > V(\sigma)$ which contradicts σ's optimality.

(ii) Let us now show that σ must always be increasing. Let (σ, τ) be any mechanism respecting (IC). Since the agents' cost has decreasing differences the function τ must be non-decreasing in θ by Topkis' theorem. Together with (IC), this implies

$$\sigma(\tau(\theta)) - \sigma(\tau(\theta')) \geq \gamma\, (c(\tau(\theta), \theta) - c(\tau(\theta'), \theta)) \geq 0,$$

for any $\theta > \theta'$. The second inequality comes frome the fact that $c(\cdot, \theta)$ is an increasing function for any θ. As a result, σ must be a non-decreasing function.

A.2.2. Proof of lemma 3

Necessity. It follows directly from proposition 2 in Rochet (1987) that property (i) is necessary and sufficient for (IC) so let us prove property (ii) first. Remark that the lowest (resp. highest) recommendation rule in the class of monotone selection rules is given by $\underline{\sigma}(t) = 0$ for all $t \in T$ (resp. $\bar{\sigma}(t) = \mathbb{1}\{t \geq 0\}$ for any $t \in T$). Hence, for any monotone selection rule σ and any $\theta \in [\theta_0, \bar{\theta}]$, we must have:

$$\underbrace{\max_{t \in T}\ t\theta - \left(\frac{t^2}{2} - \frac{\underline{\sigma}(t)}{\gamma}\right)}_{=\underline{u}(\theta)} \leq \underbrace{\max_{t \in T}\ t\theta - \left(\frac{t^2}{2} - \frac{\sigma(t)}{\gamma}\right)}_{=u(\theta)} \leq \underbrace{\max_{t \in T}\ t\theta - \left(\frac{t^2}{2} - \frac{\bar{\sigma}(t)}{\gamma}\right)}_{=\bar{u}(\theta)}.$$

Next, let us prove property (iii). If a mechanism (σ, τ) is admissible it must satisfy (IC). Hence, by property (i), we have $u'(\theta) = \tau(\theta)$ almost everywhere on $[\theta_0, \bar{\theta}]$ and corollary 2 implies directly that $\theta \leq u'(\theta) \leq \bar{t}(\theta)$ for almost every $\theta \in [\theta_0, \bar{\theta}]$. Let us now prove property (iv). If (σ, τ) is admissible, then we can again substitute $\tau(\theta)$ by $u'(\theta)$ almost everywhere on $[\theta_0, \bar{\theta}]$. Remembering that $U(\theta) = \gamma(u(\theta) - \theta^2/2)$ and remarking that $\sigma(\tau(\theta)) = U(\theta) + \gamma c(\tau(\theta), \theta)$, some algebra shows that:

$$\sigma(\tau(\theta)) = \gamma \left(u(\theta) + \frac{u'(\theta)^2}{2} - \theta u'(\theta) \right),$$

for almost every $\theta \in [\underline{\theta}, \bar{\theta}]$. Since $\sigma(t) \leq 1$ for all $t \in T$ we must have:

$$u(\theta) + \frac{u'(\theta)^2}{2} - \theta u'(\theta) \leq \frac{1}{\gamma},$$

for almost every $\theta \in [\underline{\theta}, \bar{\theta}]$. Next, we prove property (v). Since (σ, τ) is admissible, there must exist a $\theta^\dagger$ such that $u'(\theta) = \theta$ for almost all $\theta \in [\theta_0, \theta^\dagger[$ by (IC) and corollary 2. Moreover, since u is convex by (IC), it must also be absolutely continuous and thus equal to the integral of its derivative. Hence:

$$\begin{aligned} u(\theta) &= u(\theta_0) + \int_{\theta_0}^{\theta} z \, \mathrm{d}z \\ &= u(\theta_0) - \frac{\theta_0^2}{2} + \underline{u}(\theta), \end{aligned}$$

for all $\theta \in [\theta_0, \theta^\dagger[$. Let us prove that $u(\theta_0) = \theta_0^2/2$. First of all, remember that $\theta_0 = -\sqrt{2/\gamma} < 0$. Since σ is monotone, either $\sigma(0) = 1$, in which case $\tau(\theta_0) = 0$ and $u(\theta_0) = 1/\gamma = \theta_0^2/2$, or $0 \leq \sigma(0) < 1$, in which case $\tau(\theta_0) = \theta_0$ which implies that $u(\theta_0) = \theta_0^2/2$. As a convex function, u must also be continuous implying that $u(\theta^\dagger) = \underline{u}(\theta^\dagger)$. Therefore, we must have $u(\theta) = \underline{u}(\theta)$ for all $\theta \in [\theta_0, \theta^\dagger]$. We can deduce directly from (IC) and corollary 2 that $u'(\theta) \geq 0$ for almost all $\theta \in]\theta^\dagger, \bar{\theta}]$. Therefore, starting from $\theta^\dagger$, the function u must be increasing which also implies that $u(\theta) > \underline{u}(\theta)$ for all $\theta \in]\theta^\dagger, \bar{\theta}]$. Finally, let us prove property (vi). Assume that there exists $\tilde{\theta} \in [\theta_0, \bar{t}^{-1}(\bar{\theta})]$ such that $u'(\tilde{\theta}) = \bar{t}(\tilde{\theta})$. corollary 2 and (IC) imply that:

$$u'(\theta) = \begin{cases} \bar{t}(\tilde{\theta}) & \text{if } \theta \in [\tilde{\theta}, \bar{t}(\tilde{\theta})[\\ \theta & \text{if } \theta \in [\bar{t}(\tilde{\theta}), \bar{\theta}] \end{cases}.$$

Using again the absolute continuity of u, we obtain:

$$\begin{aligned} u(\theta) &= u(\tilde{\theta}) + \int_{\tilde{\theta}}^{\theta} \bar{t}(\tilde{\theta})\, \mathrm{d}z \\ &= u(\tilde{\theta}) + \bar{t}(\tilde{\theta})(\theta - \tilde{\theta}) \end{aligned} \tag{A.1}$$

for all $\theta \in [\tilde{\theta}, \bar{t}(\tilde{\theta})[$, as well as:

$$\begin{aligned} u(\theta) &= u(\bar{t}(\tilde{\theta})) + \int_{\bar{t}(\tilde{\theta})}^{\theta} z\, \mathrm{d}z \\ &= u(\bar{t}(\tilde{\theta})) - \frac{\bar{t}(\tilde{\theta})^2}{2} + \frac{\theta^2}{2} \end{aligned} \tag{A.2}$$

for all $\theta \in [\bar{t}(\tilde{\theta}), \bar{\theta}]$. Let us now prove that $u(\tilde{\theta}) = \underline{u}(\tilde{\theta})$ and $u(\bar{t}(\tilde{\theta})) = \bar{u}(\bar{t}(\tilde{\theta}))$. Assume that $u(\tilde{\theta}) > \underline{u}(\tilde{\theta})$. By equation (A.1) we have:

$$u(\bar{t}(\tilde{\theta})) = u(\tilde{\theta}) + \bar{t}(\tilde{\theta})(\bar{t}(\tilde{\theta}) - \tilde{\theta}) > \underline{u}(\tilde{\theta}) + \bar{t}(\tilde{\theta})(\bar{t}(\tilde{\theta}) - \tilde{\theta}). \tag{A.3}$$

Expanding the right hand side of the previous inequality, we obtain:

$$\underline{u}(\tilde{\theta}) + \bar{t}(\tilde{\theta})(\bar{t}(\tilde{\theta}) - \tilde{\theta}) = \frac{\tilde{\theta}^2}{2} + \tilde{\theta}\sqrt{\frac{2}{\gamma}} - \frac{2}{\gamma} = \frac{1}{\gamma} + \frac{\bar{t}(\tilde{\theta})^2}{2} = \bar{u}(\bar{t}(\tilde{\theta})). \tag{A.4}$$

Combining equations (A.3) and (A.4) leads to $u(\bar{t}(\tilde{\theta})) > \bar{u}(\bar{t}(\tilde{\theta}))$, a contradiction. Therefore, $u(\tilde{\theta}) = \underline{u}(\tilde{\theta})$. equation (A.4) therefore implies directly that $u(\bar{t}(\tilde{\theta})) = \bar{u}(\bar{t}(\tilde{\theta})) = 1/\gamma + \bar{t}(\tilde{\theta})^2/2$, which, together with equation (A.2), implies that:

$$u(\theta) = \frac{1}{\gamma} + \frac{\theta^2}{2} = \bar{u}(\theta),$$

for all $\theta \in [\bar{t}(\tilde{\theta}), \bar{\theta}]$. Finally, by continuity of u and property (v), we must have that $u(\theta) = \underline{u}(\theta)$ for all $\theta \in [\theta_0, \tilde{\theta}[$. The proof for the case where $\bar{t}(\tilde{\theta}) > \bar{\theta}$ is analogous and omitted.

Sufficiency. To prove sufficiency, we must show that if a pseudo-utility function satisfies properties (i) to (v) in lemma 3, then the mechanism (σ, τ) which induces it must be admissible. It particular, it suffices to show that σ is monotone as defined in lemma 1 since it would directly imply that τ satisfies all the properties in corollary 2. Let $u \in \mathcal{U}$ where $\mathcal{U}$ denotes the set of functions satisfying all the properties in lemma 3 and define the function $x(\theta) = \gamma(u(\theta) + u'(\theta)^2/2 - \theta u'(\theta))$

for any $\theta \in [\theta_0, \bar{\theta}]$. First, since u is convex, it must be twice differentiable almost everywhere on $[\theta_0, \bar{\theta}]$ by Alexandrov's theorem (see Aliprantis and Border, 2006, Theorem 7.28). Hence, x' exists almost everywhere on $[\theta_0, \bar{\theta}]$ and is given by:

$$\begin{aligned} x'(\theta) &= \gamma\left(u'(\theta) + u''(\theta)u'(\theta) - u'(\theta) - \theta u''(\theta)\right) \\ &= \gamma u''(\theta)(u'(\theta) - \theta) \end{aligned}$$

for almost every $\theta \in [\theta_0, \bar{\theta}]$. Hence $x'(\theta) \geq 0$ almost everywhere since $u''(\theta) \geq 0$ and $u'(\theta) \geq \theta$ almost everywhere on $[\theta_0, \bar{\theta}]$. Remarking that $x(\theta) = \sigma(\tau(\theta))$ almost everywhere on $[\theta_0, \bar{\theta}]$ and that τ must be increasing on $[\theta_0, \bar{\theta}]$ by (IC), we can conclude that the function σ must he increasing on T. Second, for any $\theta \in [\theta_0, \bar{\theta}]$ let $g(\cdot, \theta)$ be the function defined by

$$g(x, \theta) = \frac{x^2}{2} - \theta x$$

for any $x \in [\theta, \bar{t}(\theta)]$. The function $g(\cdot, \theta)$ is differentiable and

$$g_x(x, \theta) = x - \theta \geq 0$$

for any $x \in [\theta, \bar{t}(\theta)]$. Hence, $g(\cdot, \theta)$ is increasing over the interval $[\theta, \bar{t}(\theta)]$ for any $\theta \in [\theta_0, \bar{\theta}]$ and is thus minimized at θ where $g(\theta, \theta) = -\bar{u}(\theta)$. Since $u(\theta) \geq \underline{u}(\theta)$ and $u'(\theta) \geq \theta = \underline{u}'(\theta)$ we have:

$$x(\theta) = \gamma\left(u(\theta) + g(u'(\theta), \theta)\right) \geq \gamma\left(\underline{u}(\theta) + g(\underline{u}'(\theta), \theta)\right) = 0,$$

for all $\theta \in [\theta_0, \bar{\theta}]$, which implies that $\sigma(t) \geq 0$ for all $t \in T$. Moreover, property (iv) directly implies that $x(\theta) \leq 1$, implying that $\sigma(t) \leq 1$ for all $t \in T$. Finally, let us prove that $\sigma(t) = 0$ for all $t \in [\underline{\theta}, 0[$. The convexity of u also implies that the left limit of u' exists everywhere on $[\theta_0, \bar{\theta}]$. In particular, since $u(\theta_0) = \theta_0^2/2$ we have $\lim_{\theta \to \theta_0^-} u'(\theta) = \theta_0$ and therefore:

$$\begin{aligned} \lim_{\theta \to \theta_0^-} x(\theta) &= \gamma\left(\frac{\theta_0^2}{2} + \frac{\theta_0^2}{2} - \theta_0^2\right) \\ &= 0. \end{aligned}$$

which implies that $\sigma(t) = 0$ for all $t \in [\bar{\theta}, 0[$ since x is increasing and bounded below by 0.

A.2.3. Proof of lemma 4

Let us first prove convexity. Fix any $\theta^\dagger \in [\theta_0, \bar{\theta}]$ and let $u_1, u_2 \in \mathcal{U}(\theta^\dagger)$ and $\lambda \in [0,1]$. It is immediate that the function $w = \lambda u_1 + (1-\lambda)u_2$ satisfies properties (ii) and (iii). Next, remark that:

$$
\begin{aligned}
&w(\theta) + \frac{w'(\theta)^2}{2} - \theta w'(\theta) \\
&\quad \leq \lambda \left(u_1(\theta) + \frac{u_1'(\theta)^2}{2} - \theta u_1'(\theta) \right) + (1-\lambda) \left(u_2(\theta) + \frac{u_2'(\theta)^2}{2} - \theta u_2'(\theta) \right) \leq \frac{1}{\gamma}
\end{aligned}
$$

where the first inequality is obtained by applying Jensen's inequality on the term $w'(\theta)^2/2$ and the second inequality is implied by property (iv) in lemma 3. Let us now prove compactness. First, remark that $\mathcal{U}(\theta^\dagger)$ is a *closed* subspace of $\mathcal{C}([\theta^\dagger, \bar{\theta}])$ because (i) every convex function is continuous, (ii) it is defined by closed inequalities, and (iii) for any a uniformly convergent sequence $(u_n)_{n\in\mathbb{N}}$ of convex functions on the compact set $[\theta^\dagger, \bar{\theta}]$, the sequence of derivatives $(u_n')_{n\in\mathbb{N}}$ must converge uniformly to u' by Theorem 25.7 in Rockafellar (1970). Second, the set $\mathcal{U}(\theta^\dagger)$ is *uniformly bounded*: if $\theta \in [\theta^\dagger, \bar{\theta}]$ then $|u(\theta)| \leq \bar{u}(\bar{\theta})$. Third, the space $\mathcal{U}(\theta^\dagger)$ is *equicontinuous*. Indeed, if $\theta \in [\theta^\dagger, \bar{\theta}]$ and $u \in \mathcal{U}(\theta^\dagger)$ then

$$
|u'(\theta)| \leq \bar{t}(\bar{\theta}).
$$

From the Mean Value Theorem we can infer that

$$
|u(\theta) - u(\theta')| \leq \bar{t}(\bar{\theta})|\theta - \theta'|,
$$

for any $u \in \mathcal{U}(\theta^\dagger)$ and any $\theta, \theta' \in [\theta^\dagger, \bar{\theta}]$. Therefore, for any $u \in \mathcal{U}(\theta^\dagger)$, any $\theta \in [\theta^\dagger, \bar{\theta}]$ and any $\varepsilon > 0$, taking $\delta = \varepsilon/\bar{t}(\bar{\theta})$ implies

$$
|\theta - \theta'| < \delta \implies |u(\theta) - u(\theta')| < \varepsilon.
$$

Hence, by the *Arzelà-Ascoli theorem* (Royden and Fitzpatrick, 2010, Theorem 3), the space $\mathcal{U}(\theta^\dagger)$ is compact with respect to the supremum norm.

A.2.4. Proof of 5

We first prove upper semicontinuity and then convexity:

(i) Fix any $\theta^\dagger \in [\theta^\dagger, \bar{\theta}]$. The space $\mathcal{U}(\theta^\dagger)$ endowed with the distance induced by the supremum norm is a complete metric space. Therefore we can use

the sequential characterization of upper semicontinuity: The functional V is upper semicontinuous if and only if for every $u \in \mathcal{U}(\theta^\dagger)$ and $\ell \in \mathbb{R}$ we have

$$\lim_{n\to\infty} u_n = u, \lim_{n\to\infty} V_{\theta^\dagger}(u_n) \geq \ell \implies V(u) \geq \ell.$$

Let $(u_n)_{n\in\mathbb{N}}$ be an arbitrary sequence in $\mathcal{U}$ such that $(u_n)_{n\in\mathbb{N}}$ converges uniformly to some $u \in \mathcal{U}(\theta^\dagger)$ with $\lim_{n\to\infty} V(u_n) \geq \ell$. Invoking again theorem 25.7 in Rockafellar (1970) the sequence $(u'_n)_{n\in\mathbb{N}}$ must converge uniformly to u'. Since the function $\Lambda(\theta,\cdot,\cdot)$ is continuous for any $\theta \in \Theta$, we must have $\Lambda(\theta, u(\theta), u'(\theta)) = \limsup_{n\to\infty} \Lambda(\theta, u_n(\theta), u'_n(\theta))$. Fatou's (reverse) lemma in turn implies that

$$\begin{aligned} V_{\theta^\dagger}(u) &= \int_{\theta^\dagger}^{\bar\theta} \Lambda\big(\theta, u(\theta), u'(\theta)\big)\mathrm{d}\theta = \int_{\theta^\dagger}^{\bar\theta} \limsup_{n\to\infty} \Lambda\big(\theta, u_n(\theta), u'_n(\theta)\big)\mathrm{d}\theta \\ &\geq \limsup_{n\to\infty} \int_{\theta^\dagger}^{\bar\theta} \Lambda\big(\theta, u_n(\theta), u'_n(\theta)\big)\mathrm{d}\theta = \limsup_{n\to\infty} V(u_n) \geq \ell. \end{aligned}$$

(ii) We now prove that $V_{\theta^\dagger} : \mathcal{U}(\theta^\dagger) \to \mathbb{R}$ is convex provided that $f'(\theta) \leq 0$. Consider an initial pseudo-utility function $u \in \mathcal{U}(\theta^\dagger)$, and a variation $h\colon [\theta^\dagger, \bar\theta] \to \mathbb{R}$. A variation is admissible if $u + h \in \mathcal{U}(\theta^\dagger)$. Therefore, it must be that (i) $h(\theta^\dagger) = h(\bar\theta) = 0$ and that (ii) $(u+h)'$ exists almost everywhere on the interval $[\theta^\dagger, \bar\theta]$ which implies that that h' exists almost everywhere. Fix now an arbitrary $\varepsilon \in [0,1]$. Then, $u + \varepsilon h \in \mathcal{U}(\theta^\dagger)$ since $\mathcal{U}(\theta^\dagger)$ is a convex set. Consider the function $\phi\colon [0,1] \to \mathbb{R}$ defined by:

$$\phi(\varepsilon) = V_{\theta^\dagger}(u + \varepsilon v) = \int_{\theta^\dagger}^{\bar\theta} \Lambda\big(\theta, u(\theta) + \varepsilon h(\theta), u'(\theta) + \varepsilon h'(\theta)\big)\mathrm{d}\theta. \tag{A.5}$$

For any admissible variation h and any $u \in \mathcal{U}(\theta^\dagger)$, the directional derivative of $V_{\theta^\dagger}$ at u in direction h is given by $\phi'(0)$ if it exists. We say that $V_{\theta^\dagger}$ is Gâteaux differentiable at u in the direction h if and only if the directional derivative $\phi'(0)$ at u in the direction h can be written as a linear functional of h, i.e., $\phi'(0)$ is of the form $\mathrm{D}V_{\theta^\dagger}(u)(h)$ for any u and h. We then call $\mathrm{D}V_{\theta^\dagger}(u)$ the Gâteaux derivative of $V_{\theta^\dagger}$ at u. Similarly, we say that $V_{\theta^\dagger}$ is twice-Gâteaux-differentiable if $\mathrm{D}V_{\theta^\dagger}(u)(h)$ is Gâteaux differentiable at u in direction k, i.e., $\phi''(0)$ exists and can be written as a bilinear form in (h,k), denoted $\mathrm{D}^2V_{\theta^\dagger}(u)(h,k)$.

We then use the following characterization of convexity: The functional $V_{\theta^\dagger}$ is

convex if and only if $V_{\theta^\dagger}$ is twice-Gâteaux-differentiable and $D^2V_{\theta^\dagger}(u)(h,h) \geq 0$ for any admissible direction h (see for instance Clarke, 2013, Theorem 2.26). Remark that $\Lambda(\theta,\cdot,\cdot)$ is twice continuously differentiable. Hence, differentiating twice the objective under the integral sign and evaluating the expression at $\varepsilon = 0$, we obtain:

$$\begin{aligned}\phi''(0) = \int_{\theta^\dagger}^{\bar\theta} \Big(& \Lambda_{xx}\left(\theta, u(\theta), u'(\theta)\right) h'(\theta)^2 \\ & + 2\Lambda_{xy}\left(\theta, u(\theta), u'(\theta)\right) h'(\theta) h(\theta) \\ & + \Lambda_{xy}\left(\theta, u(\theta), u'(\theta)\right) h(\theta)^2 \Big) \mathrm{d}\theta.\end{aligned}$$

Integrating by parts, we then have:

$$\begin{aligned}\phi''(0) &= \int_{\theta^\dagger}^{\bar\theta} \Lambda_{yy}\left(\theta, u(\theta), u'(\theta)\right) h'(\theta)^2 \,\mathrm{d}\theta \\ &+ \int_{\theta^\dagger}^{\bar\theta} \left(\Lambda_{xx}\left(\theta, u(\theta), u'(\theta)\right) - \frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{xy}\left(\theta, u(\theta), u'(\theta)\right)\right) h(\theta)^2 \,\mathrm{d}\theta,\end{aligned}$$

since $v(\theta^\dagger) = v(\bar\theta) = 0$. This proves that $V_{\theta^\dagger}$ is twice-Gâteaux-differentiable. Replacing the terms in both integrals by their expression, we can deduce that:

$$D^2V_{\theta^\dagger}(u)(h,h) = \int_{\theta^\dagger}^{\bar\theta} \left(3u'(\theta) - 2\theta\right) f(\theta) h'(\theta)^2 \,\mathrm{d}\theta + \int_{\theta^\dagger}^{\bar\theta} \left(-f'(\theta)\right) h(\theta)^2 \,\mathrm{d}\theta$$

The second integral is always positive since $f'(\theta) \leq 0$ for any $\theta \in [\theta_0, \bar\theta]$ under assumption 1. Moreover, lemma 3 implies that $u'(\theta) \geq \theta$ as well as $u'(\theta) \geq 0$ for all $\theta \geq \theta^\dagger$. Therefore, $3u'(\theta) - 2\theta \geq 0$ for all $\theta \geq \theta^\dagger$. Hence, the first integral is also always positive, which concludes the proof.

A.2.5. Proof of lemma 6

Assume $u \in \mathcal{U}(\theta^\dagger)$ is an extreme point and that there exist $\mathcal{I} = [a,b] \subseteq \,]\theta_0, \bar\theta]$ such that $u'(\theta) > 0$, $\theta < u'(\theta) < \bar t(\theta)$ and $\underline{u}(\theta) < u(\theta) < \bar u(\theta)$ for all $\theta \in \mathcal{I}$. Then let us fix some $\varepsilon > 0$ and define the function $h_\varepsilon \colon [\theta_0, \bar\theta] \to \mathbb{R}$ such that:

$$h_\varepsilon(\theta) = \begin{cases} \dfrac{\varepsilon}{2}(\theta - a)^2(\theta - b)^2 & \text{if } \theta \in \mathcal{I} \\ 0 & \text{if } [\theta_0, \bar\theta] \setminus \mathcal{I} \end{cases},$$

The function h_ε is twice continuously differentiable and its first and second derivatives are respectively given by:

$$h'_\varepsilon(\theta) = \varepsilon(\theta - a)(\theta - b)(2\theta - (a + b)),$$

and

$$h''_\varepsilon(\theta) = \varepsilon\left((a + b)^2 + 2ab + 6(\theta^2 - (a + b)\right),$$

for all $\theta \in \mathcal{I}$. Hence, we must always have $u(\theta) \pm h_\varepsilon(\theta) = u(\theta)$ if $\theta \in [\theta_0, \bar{\theta}] \setminus \mathcal{I}$ and if $\theta \in \{a, b\}$ as well as $u'(\theta) \pm h'_\varepsilon(\theta) = u'(\theta)$ when $\theta \in \{a, b\}$. Moreover, it is always possible to choose $\varepsilon > 0$ small enough so that $\underline{u}(\theta) < u(\theta) \pm h_\varepsilon(\theta) < \bar{u}(\theta)$ and $\theta < u'(\theta) \pm h'_\varepsilon(\theta) < \bar{t}(\theta)$ for every $\theta \in \mathcal{I}$ as well as that $u''(\theta) + h''_\varepsilon(\theta) > 0$ wherever u'' exists so that $u \pm h_\varepsilon$ stays convex on $[\theta_0, \bar{\theta}]$. This implies that for a well chosen ε we must have $u \pm h_\varepsilon \in \mathcal{U}(\theta^\dagger)$, a contradiction. Therefore, if $u \in \mathcal{U}(\theta^\dagger)$ and $\underline{u}(\theta) < u(\theta) < \bar{u}(\theta)$ and $\theta < u'(\theta) < \bar{t}(\theta)$ on some interval, then u' must be constant on that interval, i.e., that u is affine on that interval. Therefore, if the function u is an extreme point, it cannot be strictly convex on an interval where neither bound on its derivative u' is binding. Accordingly, assume now that $\underline{u}(\theta) < u(\theta) < \bar{u}(\theta)$ but that $u'(\theta)$ is either equal to θ or $\bar{t}(\theta)$ on some interval. We know from lemma 3 that u' can be confounded with the bound $\bar{t}(\theta)$ at most at one point. Therefore, the only possibility for u to be strictly convex on some interval is that $u'(\theta) = \theta$ on that interval, so $u(\theta) = \theta^2/2 + c$ for some well chosen constant $0 \leq c \leq 1/\gamma$.

A.2.6. Proof of lemma 8

For any $u \in \mathcal{U}(\theta^\dagger)$ and any admissible variation h, the directional derivative of $V_{\theta^\dagger}$ at u in direction h is given by:

$$\phi'(0) = \int_{\theta^\dagger}^{\bar{\theta}} \left(\Lambda_x\big(\theta, u(\theta), u'(\theta)\big)h(\theta) + \Lambda_y\big(\theta, u(\theta), u'(\theta)\big)h'(\theta)\right) \mathrm{d}\theta \qquad \text{(A.6)}$$

where $\phi\colon [0, 1] \to \mathbb{R}$ has been defined in equation (A.5). For convenience, denote $\alpha(\theta) = \Lambda_x(\theta, u(\theta), u'(\theta))$ and $\beta(\theta) = \Lambda_y(\theta, u(\theta), u'(\theta))$. Integrating equation (A.6) by parts, we obtain:

$$\phi'(0) = \int_{\theta^\dagger}^{\bar{\theta}} (\alpha(\theta) - \beta'(\theta))\, h(\theta)\mathrm{d}\theta \qquad \text{(A.7)}$$

which proves that $V_{\theta^\dagger}$ is Gâteaux differentiable. Moreover:

$$\alpha(\theta) = u'(\theta) f(\theta), \tag{A.8}$$

and

$$\beta(\theta) = \left(\left(u(\theta) + \frac{1}{2} u'(\theta)^2 - \theta u'(\theta) \right) + u'(\theta) \left(u'(\theta) - \theta \right) \right) f(\theta), \tag{A.9}$$

for all $\theta \in [\theta^\dagger, \bar{\theta}]$. The function β defined in equation (A.9) is clearly differentiable almost everywhere on $[\theta^\dagger, \bar{\theta}]$ and has a derivative given by:

$$\begin{aligned} \beta'(\theta) = {} & \left(2u''(\theta) \left(u'(\theta) - \theta\right) + u'(\theta) \left(u''(\theta) - 1\right)\right) f(\theta) \\ & + \left(\left(u(\theta) + \frac{1}{2} u'(\theta)^2 - \theta u'(\theta) \right) + u'(\theta) \left(u'(\theta) - \theta \right) \right) f'(\theta) \end{aligned} \tag{A.10}$$

wherever it exists. Subtracting equations (A.8) and (A.10) and plugging the result in equation (A.7) yields the desired result.

A.2.7. Proof of lemma 9

First of all, we know from lemma 8 that the Gâteaux derivative of $V_{\theta^\dagger}$ has the form:

$$\begin{aligned} \mathrm{D}V_{\theta^\dagger}(u)(h) = \int_{\theta^\dagger}^{\bar{\theta}} \Bigg(& - \beta(\theta) f'(\theta) \\ & + \Big(u'(\theta) - 2u''(\theta) \left(u'(\theta) - \theta\right) - u'(\theta) \left(u''(\theta) - 1\right) \Big) f(\theta) \Bigg) h(\theta) \, \mathrm{d}\theta. \end{aligned} \tag{A.11}$$

Remark that $\beta(\theta) = x(\theta) + u'(\theta)(u'(\theta) - \theta) \geq 0$ for all $\theta \in [\theta^\dagger, \bar{\theta}]$ since $x(\theta) \geq 0$, $u'(\theta) \geq 0$ and $u'(\theta) \geq \theta$ for all $\theta \in [\theta^\dagger, \bar{\theta}]$. Moreover, under assumption 1, $-f'(\theta) \geq 0$. Therefore, the term $-\beta(\theta) f'(\theta)$ is positive for all $\theta \in [\theta^\dagger, \bar{\theta}]$. As a consequence, the sign of the integrand in equation (A.11) only depends on the second term. We are going to show that this term is always positive when $u \in \mathcal{E}(\theta^\dagger)$.

Let $u \in \mathcal{E}(\theta^\dagger)$. We know from lemma 6 that, on any subinterval of $[\theta^\dagger, 0[$, the function u must have the form $u(\theta) = a\theta + b$ for some constants $a \geq 0$ and $b \in \mathbb{R}$, and that on any subinterval of $[0, \bar{\theta}]$, the function u must either have the form $u(\theta) = a\theta + b$ or $u(\theta) = \theta^2/2 + c$ for some constants $a \geq 0$, $b \in \mathbb{R}$ and $c \in [0, 1/\gamma]$. First assume that $u(\theta) = a\theta + b$ on some interval $[\underline{x}, \bar{x}] \subseteq [\theta^\dagger, \bar{\theta}]$.

Then we must have $u''(\theta) = 0$ on $[\underline{x}, \bar{x}]$, which implies that:

$$\Big(\underbrace{-\beta(\theta) f'(\theta)}_{\geq 0} + 2\underbrace{u'(\theta)}_{\geq 0}\Big)\underbrace{\Big(u^{\dagger}(\theta) - u(\theta)\Big)}_{\geq 0} \geq 0, \tag{A.12}$$

for any $\theta \in [\underline{x}, \bar{x}]$. Now, assume that $\theta \geq 0$ and that $u(\theta) = \theta^2/2 + c$ for some constant $c \in [0, 1/\gamma]$ on some interval $[\underline{x}, \bar{x}] \subseteq [0, \bar{\theta}]$. We thus have $u'(\theta) = \theta$ and $u''(\theta) = 1$, and therefore

$$\left(\underbrace{-\beta(\theta) f'(\theta)}_{\geq 0} + \underbrace{u'(\theta)}_{\geq 0}\right)\underbrace{\Big(u^{\dagger}(\theta) - u(\theta)\Big)}_{\geq 0} \geq 0. \tag{A.13}$$

for any $\theta \in [\underline{x}, \bar{x}]$. equations (A.12) and (A.13) both imply that the integrand of the Gâteaux derivative is always positive along increasing affine arcs and increasing quadratic arcs. Hence, $\mathrm{D}V_{\theta^{\dagger}}(u)(u^{\dagger} - u) \geq 0$ for any $u \in \mathcal{E}(\theta^{\dagger})$.

A.3. Proofs for Comparative Statics

For notational ease we let $\theta^*_\gamma = \theta(\gamma)$ and $t^*_\gamma = t(\gamma)$.

A.3.1. Proof of 2

The optimal cutoff t^*_γ is solution to the equation $\psi(t) = 0$ where

$$\psi(t) = \begin{cases} t - \dfrac{F(t) - F(\theta(t))}{f(\theta(t))} & \text{if } t \in [0, \bar{\theta}[\\[2ex] t - \dfrac{1 - F(\theta(t))}{f(\theta(t))} & \text{if } t \in [\bar{\theta}, \bar{\theta} + \sqrt{2/\gamma}] \end{cases}.$$

Remark that

$$\psi(\bar{\theta}) = \begin{cases} \bar{\theta} - \dfrac{1}{f(\underline{\theta})} & \text{if } \gamma < 1/c(\bar{\theta}, \underline{\theta}) \\[2ex] \bar{\theta} - \dfrac{1 - F\Big(\bar{\theta} - \sqrt{\frac{2}{\gamma}}\Big)}{f\Big(\bar{\theta} - \sqrt{\frac{2}{\gamma}}\Big)} & \text{if } \gamma \geq 1/c(\bar{\theta}, \underline{\theta}) \end{cases}.$$

The function $\gamma \mapsto (1 - F(\bar{\theta} - \sqrt{2/\gamma}))/f(\bar{\theta} - \sqrt{2/\gamma})$ is clearly increasing in γ under assumption 1 so $\psi(\bar{\theta}) \geq \bar{\theta} - 1/f(\underline{\theta})$ for any $\gamma > 0$. Hence, if $f(\underline{\theta}) > 1/\bar{\theta}$, then $\psi(\bar{\theta}) > 0$ for all $\gamma > 0$. As in the proof of proposition 1 in appendix A.1, the Intermediate Value Theorem implies that $t^*_\gamma \in [0, \bar{\theta}]$ for any $\gamma > 0$.

A.3.2. Proof of proposition 2

Under any pass-fail mechanism, the payoff of the designer as a function of t and γ writes:

$$V(t,\gamma) = \begin{cases} t\left(F(t) - F\left(t - \sqrt{2/\gamma}\right)\right) + \int_t^{\bar{\theta}} \theta f(\theta)\,\mathrm{d}\theta & \text{if } t \in [0, \bar{\theta}[\\ t\left(1 - F\left(t - \sqrt{2/\gamma}\right)\right) & \text{if } t \in [\bar{\theta}, \bar{\theta} + \sqrt{2/\gamma}] \end{cases}.$$

For any $t \geq 0$ and $\gamma > 0$ we have:

$$V_{t\gamma}(t,\gamma) = \frac{1}{\gamma\sqrt{2\gamma}} f'\left(t - \sqrt{2/\gamma}\right) \leq 0.$$

Hence, $V(t,\gamma)$ is submodular in (t,γ). By Topkis' theorem, we can conclude that t^*_γ is a decreasing function of γ. Equivalently, we can solve for the optimal pass-fail rule by optimizing on the last approved type by operating the change of variable $\theta = t - \sqrt{2/\gamma}$. The designer's objective function in the new coordinate system (θ, γ) thus writes as follows:

$$V(\theta,\gamma) = \begin{cases} \left(\theta + \sqrt{\frac{2}{\gamma}}\right)\left(F\left(\theta + \sqrt{\frac{2}{\gamma}}\right) - F(\theta)\right) + \int_{\theta + \sqrt{\frac{2}{\gamma}}}^{\bar{\theta}} z f(z)\,\mathrm{d}z & \text{if } \theta \in [-\sqrt{2/\gamma}, \bar{\theta} - \sqrt{2/\gamma}[\\ \left(\theta + \sqrt{\frac{2}{\gamma}}\right)(1 - F(\theta)) & \text{if } \theta \in [\bar{\theta} - \sqrt{2/\gamma}, \bar{\theta}] \end{cases}.$$

Here, we have

$$V_{\theta\gamma}(\theta,\gamma) = -\frac{1}{\gamma\sqrt{2\gamma}}\Big(\underbrace{f(\theta + \sqrt{2/\gamma}) - f(\theta)}_{\leq 0}\Big) \geq 0$$

if $\theta \in [-\sqrt{2/\gamma}, \bar{\theta} - \sqrt{2/\gamma}[$, as well as

$$V_{\theta\gamma}(\theta,\gamma) = \frac{1}{\gamma\sqrt{2\gamma}} f(\theta) \geq 0$$

if $\theta \in [\bar{\theta} - \sqrt{2/\gamma}, \bar{\theta}]$. Therefore, $V(\theta, \gamma)$ is supermodular (θ, γ) and whence θ^*_γ is increasing in γ.

Finally, remember that t^*_γ must solve the equation

$$\psi(t, \gamma) = 0$$

where

$$\psi(t, \gamma) = t - \frac{F(t) - F\left(t - \sqrt{\frac{2}{\gamma}}\right)}{f\left(t - \sqrt{\frac{2}{\gamma}}\right)}.$$

Remark that we have:

$$\lim_{\gamma \to +\infty} \psi(t, \gamma) = t$$

which implies that t^*_γ and $\theta^*_\gamma = t^*_\gamma - \sqrt{2/\gamma}$ both converge to 0 as γ becomes arbitrarily large.

A.3.3. Proof of proposition 3

Designer's welfare. The designer's optimal expected payoff is given by

$$V(\gamma) = t(\gamma)\,(F(t(\gamma)) - F(\theta(\gamma))) + \int_{t(\gamma)}^{\bar{\theta}} \theta f(\theta)\,\mathrm{d}\theta.$$

Hence its derivative is given by

$$\begin{aligned} V'(\gamma) &= t'(\gamma)\,(F(t(\gamma)) - F(\theta(\gamma))) + t(\gamma)\,(t'(\gamma) f(t(\gamma)) - \theta'(\gamma) f(\theta(\gamma))) - t'(\gamma) t(\gamma) f(t(\gamma)) \\ &= \underbrace{t'(\gamma)}_{<0}\,\underbrace{(F(t(\gamma)) - F(\theta(\gamma)))}_{>0} - \underbrace{t(\gamma)}_{>0}\,\underbrace{\theta'(\gamma)}_{>0}\,\underbrace{f(\theta(\gamma))}_{>0} < 0. \end{aligned}$$

Therefore the designer's payoff is always decreasing with γ.

Agent's ex-interim welfare. The ex-interim welfare of an agent of type θ is given by

$$U(\theta, \gamma) = \begin{cases} 0 & \text{if } \theta \in [\underline{\theta}, \theta(\gamma)[\\ 1 - \gamma c(t(\gamma), \theta) & \text{if } \theta \in [\theta(\gamma), t(\gamma)[\\ 1 & \text{if } \theta \in [t(\gamma), \bar{\theta}] \end{cases}$$

A marginal change in γ only has an effect for agents whose types are in the interval $\theta \in [\theta(\gamma), t(\gamma)[$ where we have

$$U_\gamma(\theta,\gamma) = \underbrace{-c(t(\gamma),\theta)}_{\text{Direct effect}<0} \underbrace{-\gamma t'(\gamma) c_t(t(\gamma),\theta)}_{\text{Indirect effect}>0}.$$

Given the quadratic form of the cost function we have

$$U_\gamma(\theta,\gamma) = \frac{1}{2}(t(\gamma)-\theta)(\theta - t(\gamma) - 2\gamma t'(\gamma))$$

and hence $U_\gamma(\theta,\gamma) \geq 0$ if and only if $\theta \geq t(\gamma) + 2\gamma t'(\gamma) \triangleq \tilde{\theta}(\gamma)$. Since $t'(\gamma) < 0$ we have $\tilde{\theta}(\gamma) < t(\gamma)$. Moreover, since $t(\gamma) = \theta(\gamma) + \sqrt{2/\gamma}$ we have $\tilde{\theta}(\gamma) = \theta(\gamma) + 2\gamma\theta'(\gamma) > \theta(\gamma)$ because $\theta'(\gamma) > 0$. Therefore, there exists a threshold $\tilde{\theta}(\gamma) \in]\theta(\gamma), t(\gamma)[$ above which $U'_\gamma(\theta) \geq 0$ and below which $U'_\gamma(\theta) \leq 0$.

A.4. Proof of proposition 5

First of all, (IC) implies that σ must be non-decreasing on T (see the proof of lemma 1 in appendix A.2.1). This implies that $\tau(\theta) \geq \theta$ for any $\theta \in \Theta$ under the welfare-optimal selection rule. Moreover, it is still true that $\tau(\theta) \leq \bar{t}(\theta)$ for all $\theta \in \Theta$. Remark also that any implementable pseudo-utility function must be bounded below by $\underline{u}(\theta) = \theta^2/2$ and above by $\bar{u}(\theta) = 1/\gamma + \theta^2/2$ for all $\theta \in \Theta$. The lower bound corresponds to the pseudo-utility under the selection rule $\underline{\sigma}(t) = 0$ for all $t \in T$, while the upper bound corresponds to the case where the allocation is given by $\bar{\sigma}(t) = 1$ for any $t \in T$. We call *admissible* any mechanism such that $\sigma\colon T \to [0,1]$ is non-decreasing, and that $\tau\colon \Theta \to T$ is non decreasing and bounded in between θ and $\bar{t}(\theta)$. We have the following characterization of admissible mechanisms.

Lemma 14. *A mechanism (σ,τ) is admissible if, and only if, the pseudo-utility u implemented by σ satisfies the following properties:*

(i) The function u is convex over Θ. As a result, u is differentiable a.e. on Θ and satisfies the envelope condition $u'(\theta) = \tau(\theta)$ wherever it is differentiable.

(ii) $\underline{u}(\theta) \leq u(\theta) \leq \bar{u}(\theta)$ for any $\theta \in \Theta$.

(iii) $\theta \leq u'(\theta) \leq \bar{t}(\theta)$ for any $\theta \in \Theta$.

Using the characterization from lemma 14, we can recast the program of the planner as follows:

$$\max_{u \in \mathcal{U}} \int_{\underline{\theta}}^{\bar{\theta}} \left(u'(\theta) \left(u(\theta) + \frac{u'(\theta)^2}{2} - \theta u'(\theta) \right) + \alpha \left(u(\theta) - \frac{\theta^2}{2} \right) \right) f(\theta) \, \mathrm{d}\theta.$$

where $\mathcal{U} = \{u \in \mathcal{C}(\Theta) \,|\, u \text{ convex}, \underline{u}(\theta) \leq u(\theta) \leq \bar{u}(\theta), \theta \leq u'(\theta) \leq \bar{t}(\theta)\}$. Remark that the variational program of the planner is identical to the one of the principal up to an additional linear term. The rest of the proof thus follows the same methodology as for theorem 1 and details are therefore omitted. First, we parametrize the problem with respect to the first initial type $\theta^\dagger$ investing in a positive final type. Then, we prove that the set of convex and increasing functions on the interval $[\theta^\dagger, \bar{\theta}]$ such that $u(\theta^\dagger) = \underline{u}(\theta^\dagger)$ is compact and convex for any $\theta^\dagger \in \Theta$. It is not hard to show that the objective functional admits the exact same second-order Gâteaux derivative than the objective function of the principal. Therefore, it is also convex under assumption 1. Finally, the necessary conditions on extreme points are identical to lemma 6 and the tangent inequality applies in the same way since the Gâteaux derivative is identical up to an additive term in α.

Now, let us see what is the optimal allocation cutoff. Under a pass-fail rule with cutoff t, the welfare is given by:

$$W(t) = t\big(F(t) - F(\theta(t))\big) + \int_{\theta(t)}^{t} \theta f(\theta) \, \mathrm{d}\theta + \alpha \left(1 - F(\theta(t)) - \int_{\theta(t)}^{t} \gamma c(t, \theta) f(\theta) \, \mathrm{d}\theta \right).$$

Whenever the first-order condition $W'(t) = 0$ has a solution, it is given by the solution to the equation:

$$t - \Big(1 - \alpha\gamma\big(t - \mathbb{E}[\theta \,|\, \theta(t) \leq \theta \leq t]\big) \Big) \frac{F(t) - F(\theta(t))}{f(\theta(t))} = 0. \qquad \text{(FOCW)}$$

The previous equation is identical to (FOC') up to the multiplicative term $1 - \alpha\gamma(t - [\theta \,|\, \theta(t) \leq \theta \leq t])$. Since, $\alpha\gamma \geq 0$ and $t - [\theta \,|\, \theta(t) \leq \theta \leq t] \geq 0$, we must have $1 - \alpha\gamma(t - [\theta \,|\, \theta(t) \leq \theta \leq t]) \leq 1$. As a result, whenever it exists, the solution to (FOCW) must be lower than the solution to (FOC').

A.5. Proof of lemma 10

Let's first prove (i). First, let τ be (A-IC) under σ. Note that:

$$\hat{t}_{\varsigma,\tau}(1) = \frac{\displaystyle\int_{S\times T\times\Theta} \hat{t}_{\sigma,\tau}(s)\alpha_{\sigma,\tau}(s)\,\sigma(\mathrm{d}s\,|\,t)\,\tau(\mathrm{d}t\,|\,\theta)\,\pi(\mathrm{d}\theta)}{\displaystyle\int_{S\times T\times\Theta} \alpha_{\sigma,\tau}(s)\,\sigma(\mathrm{d}s\,|\,t)\,\tau(\mathrm{d}t\,|\,\theta)\,\pi(\mathrm{d}\theta)} \geq 0,$$

and

$$\hat{t}_{\varsigma,\tau}(0) = \frac{\displaystyle\int_{S\times T\times\Theta} \hat{t}_{\sigma,\tau}(s)(1-\alpha_{\sigma,\tau}(s))\,\sigma(\mathrm{d}s\,|\,t)\,\tau(\mathrm{d}t\,|\,\theta)\,\pi(\mathrm{d}\theta)}{\displaystyle\int_{S\times T\times\Theta} (1-\alpha_{\sigma,\tau}(s))\,\sigma(\mathrm{d}s\,|\,t)\,\tau(\mathrm{d}t\,|\,\theta)\,\pi(\mathrm{d}\theta)} \leq 0,$$

since, by definition, $\alpha_{\sigma,\tau}$ and τ are best responses to each other under (σ, S). Hence, it is individually rational to follows action recommendations for the designer. That is, $\alpha_{\varsigma,\tau}(1) = 1$ and $\alpha_{\varsigma,\tau}(0) = 0$. To prove (ii), note that

$$\begin{aligned}\rho_{\varsigma,\tau}(\tau',\theta) &= \int_T \alpha_{\varsigma,\tau}(1)\,\varsigma(1\,|\,t)\,\tau'(\mathrm{d}t\,|\,\theta)\\ &= \int_{S\times T} \alpha_{\sigma,\tau}(s)\,\sigma(\mathrm{d}s\,|\,t)\,\tau'(\mathrm{d}t\,|\,\theta) = \rho_{\sigma,\tau}(\tau',\theta),\end{aligned}$$

since $\alpha_{\varsigma,\tau}(1) = 1$ by (i), and $\varsigma(1\,|\,t) = \sigma(\{s \in S : \alpha_{\sigma,\tau}(s) = 1\}|t)$ for every t. Using the fact that τ is (A-IC) under (σ, S) and that interim approval probabilities are unchanged we have that

$$\begin{aligned}\rho_{\varsigma,\tau}(\tau',\theta) - C(\tau',\theta) &= \rho_{\sigma,\tau}(\tau',\theta) - C(\tau',\theta)\\ &\leq \rho_{\sigma,\tau}(\tau,\theta) - C(\tau,\theta) = \rho_{\varsigma,\tau}(\tau,\theta) - C(\tau,\theta)\end{aligned}$$

for every type θ and investment strategies τ', which proves (iii). Finally, combining (ii) and (iii) proves (iv).

B. Mathematical appendix for chapter 2

B.1. Proof for proposition 7

Let Θ be any Polish space and let $\Delta(\Theta)$ be the set of probability measures on Θ endowed with its Borel σ-algebra, let also $\mathcal{C}_b(\Theta)$ be the set of bounded continuous and Borel-measurable real-valued functions on Θ.

For any $\eta, \mu \in \Delta(\Theta)$, by application of the Donsker-Varadhan variational formula (see Dupuis and Ellis, 1997, Lemma 1.4.3) we have

$$C(\eta,\mu) = \sup_{u(a,\cdot)\in\mathcal{C}_b(\Theta)} \int_\Theta \rho u(a,\theta)\,\eta(\mathrm{d}\theta) - \ln\left(\int_\Theta \exp\left(\rho u(a,\theta)\right)\,\mu(\mathrm{d}\theta)\right). \quad \text{(B.1)}$$

Taking the Legendre-Fenchel's dual to the variational equality (B.1) (see Dupuis and Ellis, 1997, Proposition 1.4.2) we get

$$\ln\left(\int_\Theta \exp\left(\rho u(a,\theta)\right)\,\mu(\mathrm{d}\theta)\right) = \sup_{\eta\in\Delta(\Theta)} \int_\Theta \rho u(a,\theta)\,\eta(\mathrm{d}\theta) - C(\eta,\mu). \quad \text{(B.2)}$$

Hence, we have

$$\Psi_a(\mu) = \frac{1}{\rho}\ln\left(\int_\Theta \exp\left(\rho u(a,\theta)\right)\,\mu(\mathrm{d}\theta)\right),$$

for any $a \in A$, any $\mu \in \Delta(\Theta)$ and any $\rho \in \mathbb{R}^*_+$. Moreover, the supremum in equation (B.2) is attained uniquely by the probability measure $\eta_a(\mu) \in \Delta(\Theta)$ defined by

$$\eta_a(\mu)(\tilde{\Theta}) = \frac{\int_{\tilde{\Theta}} \exp\left(\rho u(a,\theta)\right)\,\mu(\mathrm{d}\theta)}{\int_\Theta \exp\left(\rho u(a,\theta)\right)\,\mu(\mathrm{d}\theta)},$$

for any Borel set $\tilde{\Theta}$ (see, again, Dupuis and Ellis, 1997, Proposition 1.4.2).

B.2. Overoptimism about preferred outcomes

Fix an $a \in A$ and let Θ_a be the (measurable) set of states such that $\Theta_a = \arg\max_{\theta\in\Theta} u(a,\theta)$. Define $\delta(a,\theta) = u(a,\theta) - u(a,\theta^*)$ for all θ and some

$\theta^* \in \Theta_a$. Remark that $\eta_a(\mu)(\Theta_a)$ can be expressed as follows:

$$\begin{aligned}
\eta_a(\mu)(\Theta_a) &= \frac{\displaystyle\int_{\Theta_a} \exp\left(\rho u(a,\theta)\right)\, \mu(\mathrm{d}\theta)}{\displaystyle\int_{\Theta} \exp\left(\rho u(a,\theta)\right)\, \mu(\mathrm{d}\theta)} \\
&= \frac{\mu(\Theta_a)}{\mu(\Theta_a) + \displaystyle\int_{\Theta\setminus\Theta_a} \exp(\rho\delta(a,\theta))\, \mu(\mathrm{d}\theta)}.
\end{aligned}$$

Let's define the function

$$h(\rho) = \frac{\mu(\Theta_a)}{\mu(\Theta_a) + \displaystyle\int_{\Theta\setminus\Theta_a} \exp\left(\rho\delta(a,\theta)\right)\, \mu(\mathrm{d}\theta)}$$

for any $\rho \in \mathbb{R}_+^*$.

First, remark that $h(0) = \mu(\Theta_a)$. Moreover, by Leibniz integral rule, we have

$$h'(\rho) = \frac{-\mu(\Theta_a)}{\displaystyle\int_{\Theta\setminus\Theta_a} \delta(a,\theta)\exp\left(\rho\delta(a,\theta)\right)\, \mu(\mathrm{d}\theta)} \geq 0$$

for any $\rho \in \mathbb{R}_+^*$, since $\delta(a,\theta) \leq 0$. Finally, we also have that $\lim_{\rho\to+\infty} h(\rho) = 1$. Hence the probability of payoff maximizing states is bounded below by the Bayesian posterior $\mu(\Theta_a)$, is always increasing and is converging to 1 from below. Hence, a wishful Receiver always puts more probability mass on Θ_a than a Bayesian and eventually believes that the state belongs to Θ_a with probability 1 when ρ becomes large.

B.3. Proof for lemma 11

Let us study the properties of the belief threshold μ^W as a function of ρ and payoffs. First of all, let us define the function

$$\mu^W(\rho) = \frac{\exp(\rho\underline{u}_0) - \exp(\rho\underline{u}_1)}{\exp(\rho\underline{u}_0) - \exp(\rho\underline{u}_1) + \exp(\rho\overline{u}_1) - \exp(\rho\overline{u}_0)}.$$

for any $\rho \in \mathbb{R}_+^*$. To avoid notational burden, we omit the superscript W in the proof. We can find the limit of $\mu(\rho)$ at 0 by applying l'Hôpital's rule

$$\begin{aligned}\lim_{\rho\to 0}\mu(\rho) &= \lim_{\rho\to 0}\frac{\underline{u}_0\exp(\rho\underline{u}_0) - \underline{u}_1\exp(\rho\underline{u}_1)}{\underline{u}_0\exp(\rho\underline{u}_0) - \underline{u}_1\exp(\rho\underline{u}_1) + \overline{u}_1\exp(\rho\overline{u}_1) - \overline{u}_0\exp(\rho\overline{u}_0)}\\ &= \frac{\underline{u}_0 - \underline{u}_1}{\underline{u}_0 - \underline{u}_1 + \overline{u}_1 - \overline{u}_0}\\ &= \mu^B.\end{aligned}$$

So, we are back to the case of a Bayesian Receiver whenever the cost of distortion becomes infinitely high. After multiplying by $\exp(-\rho\underline{u}_0)$ at the numerator and the denominator of $\mu(\rho)$ we get

$$\mu(\rho) = \frac{1 - \exp(\rho(\underline{u}_1 - \underline{u}_0))}{1 - \exp(\rho(\underline{u}_1 - \underline{u}_0)) + \exp(\rho(\overline{u}_1 - \underline{u}_0)) - \exp(\rho(\overline{u}_0 - \underline{u}_0))}.$$

So the limit of μ^W at infinity only depends on the sign of $\overline{u}_1 - \underline{u}_0$ as, by assumption, $\underline{u}_1 - \underline{u}_0 < 0$ and $\overline{u}_0 - \underline{u}_0 < 0$. Hence, $\lim_{\rho\to+\infty}\mu(\rho) = 1$ when $\overline{u}_1 - \underline{u}_0 < 0$ and $\lim_{\rho\to+\infty}\mu(\rho) = 0$ when $\overline{u}_1 - \underline{u}_0 > 0$. Finally, in the case where $\underline{u}_0 = \overline{u}_1$ we have

$$\begin{aligned}\lim_{\rho\to+\infty}\mu(\rho) &= \lim_{\rho\to+\infty}\frac{1 - \exp(\rho(\underline{u}_1 - \underline{u}_0))}{2 - \exp(\rho(\underline{u}_1 - \underline{u}_0)) - \exp(\rho(\overline{u}_0 - \underline{u}_0))}\\ &= \frac{1}{2}.\end{aligned}$$

Let us now check the variations of the function. After differentiating with respect to ρ and rearranging terms, one can remark that the derivative of $\mu(\rho)$ must verify the following logistic differential equation with varying coefficient

$$\mu'(\rho) = \alpha(\rho)\mu(\rho)(1 - \mu(\rho)),$$

where

$$\alpha(\rho) = \frac{\underline{u}_0\exp(\rho\underline{u}_0) - \underline{u}_1\exp(\rho\underline{u}_1)}{\exp(\rho\underline{u}_0) - \exp(\rho\underline{u}_1)} - \frac{\overline{u}_1\exp(\rho\overline{u}_1) - \overline{u}_0\exp(\rho\overline{u}_0)}{\exp(\rho\overline{u}_1) - \exp(\rho\overline{u}_0)},$$

for all $\rho \in \mathbb{R}_+^*$, together with the initial condition $\mu(0) = \mu^B$. Hence, α completely dictates the variations of $\mu(\rho)$. Let us study the properties of the function α defined

on $\mathbb{R}_+^*$. First, still applying again l'Hôpital's rule, its limits are given by

$$\begin{aligned}\lim_{\rho\to 0}\alpha(\rho) &= \frac{\underline{u}_0-\overline{u}_0-(\overline{u}_1-\underline{u}_1)}{2}\\ &= \frac{1}{2}(u_0-u_1)\end{aligned}$$

and

$$\begin{aligned}\lim_{\rho\to +\infty}\alpha(\rho) &= \underline{u}_0-\overline{u}_1\\ &= u_{\max}.\end{aligned}$$

Second, after rearranging terms, its derivative is given by

$$\alpha'(\rho) = \frac{(\underline{u}_0-\underline{u}_1)^2}{\cosh(\rho(\underline{u}_0-\underline{u}_1))-1} - \frac{(\overline{u}_1-\overline{u}_0)^2}{\cosh(\rho(\overline{u}_1-\overline{u}_0))-1},$$

for any $\rho \in \mathbb{R}_+^*$, where cosh is the hyperbolic cosine function defined by

$$\cosh(x) = \frac{e^x+e^{-x}}{2},$$

for any $x \in \mathbb{R}$. Remark that the function defined by

$$f(x) = \frac{x^2}{\cosh(\rho x)-1} \tag{B.3}$$

is strictly decreasing on $\mathbb{R}_+^*$. So, we have $\alpha'(\rho) < 0$ and therefore μ^W strictly decreasing for all $\rho \in \mathbb{R}_+^*$ if and only if $\underline{u}_0-\underline{u}_1 > \overline{u}_1-\overline{u}_0$. Accordingly, α is always a strictly monotonic function if and only if $\underline{u}_0 \neq \overline{u}_1$ and $\overline{u}_0 \neq \underline{u}_1$. Hence, excluding the extreme case where $\underline{u}_0 = \overline{u}_1$ and $\overline{u}_0 = \underline{u}_1$ so $\alpha'(\rho) = 0$ and $\mu(\rho) = \mu^B$ for all $\rho \in \mathbb{R}_+^*$, three interesting cases arise, all depicted on figure B.1 for different payoff matrices:

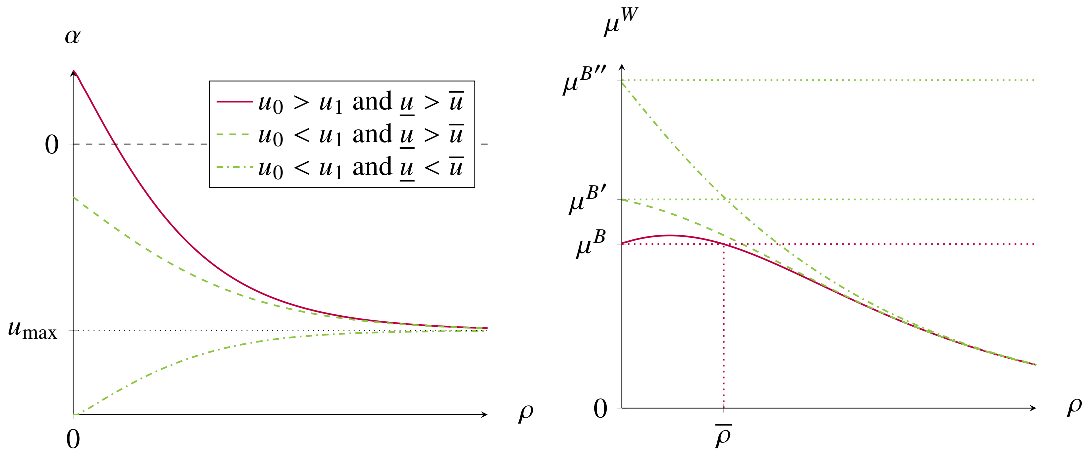
(i) If $u_{\max} < 0$, function α has a constant sign for any $\rho \in \mathbb{R}_+^*$ if and only if $u_0 < u_1$, in which case μ^W is strictly decreasing from μ^B to 0. In case $u_0 > u_1$, α has a varying sign so μ^W starts from μ^B and is sequentially strictly increasing and strictly decreasing toward 0.

(ii) If $u_{\max} = 0$, function α has a constant sign for any $\rho \in \mathbb{R}_+^*$. In this case μ^W is strictly increasing from μ^B to $1/2$ if and only if $u_0 > u_1$.

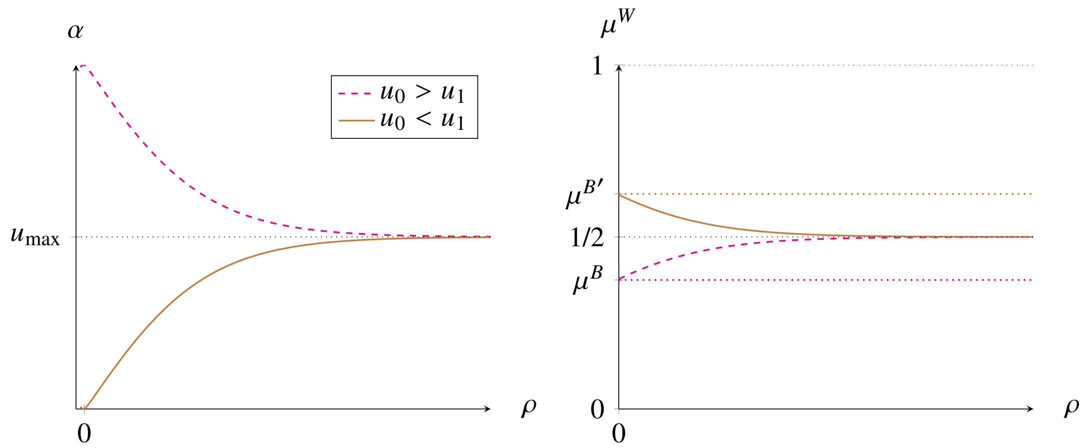
(iii) If $u_{\max} > 0$, function α has a constant sign for any $\rho \in \mathbb{R}_+^*$ if and only

if $u_0 > u_1$, in which case μ^W is strictly increasing from μ^B to 1. In case $u_0 < u_1$, α has a varying sign so μ^W starts from μ^B and is sequentially strictly decreasing and strictly increasing toward 1.

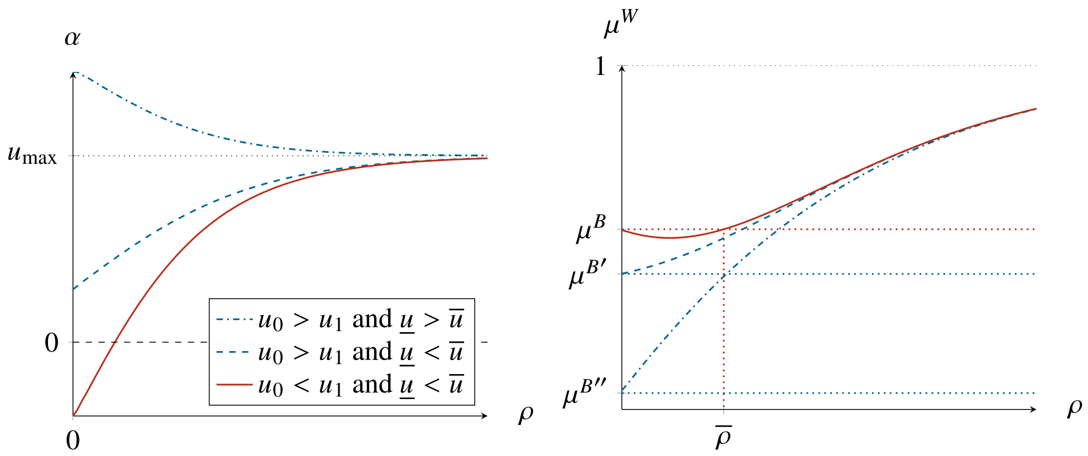
Accordingly, in case μ^W is non-monotonic in ρ, there always exists some $\overline{\rho} > 0$ such that $\mu^W(\overline{\rho}) = \mu^B$. This concludes the proof.



(a) Functions α and μ^W when $u_{\max} < 0$.



(b) Functions α and μ^W when $u_{\max} = 0$.



(c) Functions α and μ^W when $u_{\max} > 0$.

Figure B.1: Functions α and μ^W for different payoff matrices $(u_a^\theta)_{a,\theta \in A \times \Theta}$. Action $a = 1$ is favored by a wishful Receiver whenever $\mu^W < \mu^B$.

B.4. Proof for proposition 8

Assume $|\Theta| = n$ where $2 \leq n < \infty$. We want to show that $\Delta_1^B \subset \Delta_1^W$ if, and only if, the payoff matrix $(u(a,\theta))_{(a,\theta)\in A\times\Theta}$ and the wishfulness ρ verify at least one of property (i), (ii) or (iii) in lemma 11 for every pair of states $\theta, \theta' \in \Theta$.

Extreme point representation for Δ_1^B and Δ_1^W. First, remark that Δ_a^B and Δ_a^W are both convex polytopes in $\mathbb{R}^{|\Theta|}$ defined by

$$\Delta_a^B = \Delta(\Theta) \cap \left\{\mu \in \mathbb{R}^{|\Theta|} \mid \sum_{\theta\in\Theta} u(a,\theta)\mu(\theta) \geq \sum_{\theta\in\Theta} u(a',\theta)\mu(\theta),\ \forall a' \in A\right\},$$

and

$$\begin{aligned}\Delta_a^W &= \Delta(\Theta) \\ &\cap \left\{\mu \in \mathbb{R}^{|\Theta|} \mid \sum_{\theta\in\Theta} \exp\left(\rho u(a,\theta)\right)\mu(\theta) \geq \sum_{\theta\in\Theta} \exp\left(\rho u(a',\theta)\right)\mu(\theta),\ \forall a' \in A\right\}.\end{aligned}$$

The sets Δ_a^B and Δ_a^B are thus compact and convex sets in $\mathbb{R}^{|\Theta|}$ with finitely many extreme points. Let us now characterize the sets of extreme points of Δ_1^B and Δ_1^W. For any $\mu \in \mathbb{R}^{|\Theta|}$, define the systems of equations

$$\mathbf{A}^B \cdot \mu = \mathbf{b}, \quad \mu \geq 0$$

and

$$\mathbf{A}^W \cdot \mu = \mathbf{b}, \quad \mu \geq 0$$

where

$$\mathbf{A}^B = \begin{pmatrix} u^B(\theta_1) & \dots & u^B(\theta_n) \\ 1 & \dots & 1 \end{pmatrix},$$

and

$$\mathbf{A}^B = \begin{pmatrix} u^W(\theta_1) & \dots & u^W(\theta_n) \\ 1 & \dots & 1 \end{pmatrix},$$

are $2 \times n$ matrices, where $u^B(\theta) = u(1,\theta) - u(0,\theta)$ and $u^W(\theta) = \exp(\rho u(1,\theta)) - \exp(\rho u(0,\theta))$ for any $\theta \in \Theta$, and

$$\mathbf{b} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

In what follows, we always assume that $(u^B(\theta))_{\theta\in\Theta}$ and $(u^W(\theta))_{\theta\in\Theta}$ are such that $\mathrm{rank}(\mathbf{A}^B) = \mathrm{rank}(\mathbf{A}^W) = 2$.[1] Let us recall some mathematical preliminaries.

Definition 10 (Basic feasible solution)**.** *Let $\theta, \theta' \in \Theta$ be any pair of states. A vector μ^* is a basic feasible solution to $\mathbf{A}^B \cdot \mu = \mathbf{b}$ (resp. $\mathbf{A}^W \cdot \mu = \mathbf{b}$), $\mu \geq 0$, for θ, θ' if $\mathbf{A}^B \cdot \mu^* = \mathbf{b}$ (resp. $\mathbf{A}^W \cdot \mu = \mathbf{b}$), $\mu^*(\theta), \mu^*(\theta') > 0$ and $\mu^*(\theta'') = 0$ for any $\theta'' \neq \theta, \theta'$.*

Lemma 15 (Extreme point representation for convex polyhedra)**.** *A vector $\mu \in \mathbb{R}^{|\Theta|}$ is an extreme point of the convex polyhedron Δ_1^B (resp. Δ_1^B) if, and only if μ is a basic feasible solution to $\mathbf{A}^B \cdot \mu = \mathbf{b}$, $\mu \geq 0$ (resp. $\mathbf{A}^W \cdot \mu = \mathbf{b}$, $\mu \geq 0$).*

Proof. See Panik (1993) Theorem 8.4.1. □

Therefore, to find extreme points of Δ_1^B, we just have to solve the system of equations

$$\begin{cases} \mu(\theta)u^B(\theta) + \mu(\theta')b(\theta') = 0 \\ \mu(\theta) + \mu(\theta') = 1 \\ \mu(\theta), \mu(\theta') \geq 0 \end{cases} \tag{B.4}$$

for any pair of states θ, θ'. When either $\mu(\theta) = 0$ or $\mu(\theta') = 0$, the solution to (B.4) is given by the Dirac measure δ_θ only if $u^B(\theta) \geq 0$. Denote $\mathcal{E}_1^B$ the set of such beliefs. The set $\mathcal{E}_1^B$ then corresponds to the set of degenerate beliefs under which a Bayesian Receiver would take action $a = 1$. Now, if $\mu(\theta), \mu(\theta') > 0$ then the solution to (B.4) is given by

$$\mu^B_{\theta,\theta'} = \frac{u(0,\theta') - u(1,\theta')}{u(0,\theta') - u(1,\theta') + u(0,\theta) - u(1,\theta)}.$$

Such a belief is exactly the belief on the edge of the simplex between δ_θ and $\delta_{\theta'}$ at which a Bayesian decision-maker is indifferent between action $a = 0$ and $a = 1$. Denote $\mathcal{I}^B$ the set of such beliefs. Hence, we have

$$\mathrm{ext}(\Delta_1^B) = \mathcal{E}_1^B \cup \mathcal{I}^B.$$

Following the same procedure, the set of extreme points of Δ_1^W is given by $\mathcal{E}_1^W \cup \mathcal{I}^W$, where $\mathcal{E}_1^W$ is the set of degenerate beliefs at which $u^W(\theta) \geq 0$ and $\mathcal{I}^W$ is the set of beliefs

$$\mu^W_{\theta,\theta'}(\rho) = \frac{\exp(\rho u(0,\theta')) - \exp(\rho u(1,\theta'))}{\exp(\rho u(0,\theta')) - \exp(\rho u(1,\theta')) + \exp(\rho u(0,\theta)) - \exp(\rho u(1,\theta))},$$

[1]This amounts to assuming that payoff are not constant across states.

for any $\theta, \theta' \in \Theta$. Now, applying Krein-Milman theorem, we can state that

$$\Delta_1^B = \text{co}\left(\mathcal{E}_1^B \cup \mathcal{I}^B\right)$$

and

$$\Delta_1^W = \text{co}\left(\mathcal{E}_1^W \cup \mathcal{I}^W\right)$$

Sufficiency. Assume the payoff matrix $(u(a,\theta))_{(a,\theta)\in A\times\Theta}$ and the wishfulness ρ verify at least one of property (i), (ii) or (iii) in lemma 11 for every pair of states $\theta, \theta' \in \Theta$. Therefore, we have $\mu^W_{\theta,\theta'}(\rho) > \mu^B_{\theta,\theta'}$ for any $\theta, \theta' \in \Theta$. This implies $\mathcal{I}_1^B \subset \Delta_1^W$, since action $a = 1$ is favored by a wishful Receiver on each edge of the simplex. Moreover, it is trivially satisfied that $\mathcal{E}_1^B = \mathcal{E}_1^W$. Hence, since any point in Δ_1^B can be written as a convex combination of points in $\mathcal{E}_1^B \cup \mathcal{I}^B \subset \Delta^W$, it follows that $\Delta_1^B \subset \Delta_1^W$.

Necessity. Assume now that $\Delta_1^B \subset \Delta_1^W$. Therefore, we have $\mu^W_{\theta,\theta'}(\rho) > \mu^B_{\theta,\theta'}$ for any $\theta, \theta' \in \Theta$ which implies that $(u(a,\theta))_{(a,\theta)\in A\times\Theta}$ and the wishfulness ρ verify at least one of property (i), (ii) or (iii) in lemma 11 for every pair of states $\theta, \theta' \in \Theta$.

B.5. Proof for proposition 11

First, note that we can always index the voters in an ascending order of β, such that $\eta(\mu, \beta^i) \geq \eta_j(\mu)$ for all $\mu \in \Delta(\Theta)$ whenever $i < j$, such that

$$\pi(\mu) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \eta(\mu, \beta^i) - \eta(\mu, \beta^j)$$

does indeed represent the absolute difference between each pair of beliefs. Now, remark that the sum can be rearranged in the following way:

$$\begin{aligned}
\pi(\mu) =& \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \eta(\mu, \beta^i) - \eta(\mu, \beta^j) \\
=&(n-1)\eta^1(\mu) + (n-2)\eta^2(\mu) - \eta^2(\mu) + \\
& \cdots + \frac{n-1}{2}\eta(\mu, \beta^m) - \frac{n-1}{2}\eta(\mu, \beta^m) + \cdots + \\
& \eta(\mu, \beta^{n-1}) - (n-2)\eta(\mu, \beta^{n-1}) - (n-1)\eta^n(\mu) \\
=& \sum_{i=1}^{m}(n+1-2i)(\eta(\mu, \beta^i) - \eta(\mu, \beta^{n+1-i})),
\end{aligned}$$

for any $\mu \in [0, 1]$, where $m = (n+1)/2$. That is, we can express it in terms of the differences in beliefs among voters who are equidistant from the median. To see that this is true, we need to first realize that each belief appears $n-1$ times in equation (2.6) (since each belief is paired once with each of the other $n-1$ beliefs). The beliefs of voters below the median appear more often as positive than negative (the belief of the first voter is positive in all of its pairings, the belief of the second voter is positive in all of its pairing except for the pairing with the first voter, etc.), whereas the beliefs of voters above the median are more often negative than positive. If we rearrange the terms of the sum in order to pair symmetric voters, the term $(\eta(\mu, \beta^1) - \eta_n(\mu))$ appears $n-1$ times, whereas the term $(\eta_2(\mu) - \eta(\mu, \beta^{n-1}))$ appears $n-3$ times, since out of the $n-1$ times $\eta_2(\mu)$ appears on equation (2.6), $n-2$ of them are positive and 1 is negative (the converse is true for $\eta(\mu, \beta^{n-1})$). One can continue the same reasoning for all the pairs of symmetric voters, and get to the formulation of $\pi(\mu)$ presented above. Note, also, that the belief of the median voter is summed and subtracted at the same rate, such that it does not matter in our measure of polarization.

Consider the distance between beliefs of any pair of symmetric voters $\eta(\mu, \beta^i) - \eta(\mu, \beta^{n+1-i})$ for $i \in \{1, \ldots, m\}$. Given our symmetry assumption these two agents are symmetric, such that $\beta^i = 1 - \beta^{n+1-i}$. It is not difficult to show that any of those pairwise distances is maximized when agent i is distorting its belief upwards and agent $n+1-i$ is distorting its belief downwards. That is, when $\mu \in [\mu^W(\beta^i), \mu^W(\beta^{n+1-i})]$.

First, the distance between symmetric beliefs in such an interval can be rewritten

as

$$\eta(\mu, \beta^i) - \eta(\mu, \beta^{n+1-i}) = \frac{\mu \exp(\rho\beta^i)}{\mu \exp(\rho\beta^i) + (1-\mu)} - \frac{\mu}{\mu + (1-\mu)\exp(\rho\beta^i)}.$$

for any $i \in \{1, \ldots, m\}$ and $\mu \in [\mu^W(\beta^i), \mu^W(\beta^{n+1-i})]$.

Second, by taking the first order condition in this interval and rearranging it we get

$$\frac{\mu + (1-\mu)\exp(\rho\beta^i)}{\mu \exp(\rho\beta^i) + (1-\mu)} = 1,$$

such that the difference between symmetric beliefs is maximized uniquely at

$$\mu = \mu^W(\beta^m) = \frac{1}{2},$$

for any $i \in \{1, \ldots, m\}$, $\beta^i \in \,]0, 1[$ and any $\rho \in \mathbb{R}_+^*$. Since

$$\mu^W(\beta^m) = \arg\max_{\mu\in[0,1]} \eta(\mu, \beta^i) - \eta(\mu, \beta^{n+1-i})$$

for any $i \in \{1, \ldots, m\}$, we get

$$\mu^W(\beta^m) = \arg\max_{\mu\in[0,1]} \pi(\mu),$$

which concludes the proof.

B.6. Proof for proposition 10

First, we define the function

$$\psi(z) = \frac{1}{1-F(z)} \int_z^{\overline{\theta}} \exp(\rho\theta) f(\theta)\,\mathrm{d}\theta,$$

for any $z \in [\underline{\theta}, \overline{\theta}[$ and adopt the convention that $\psi(\overline{\theta}) = \exp(\rho\overline{\theta})$. It is not difficult to show that ψ is a continuous and strictly increasing function from $\psi(\underline{\theta}) = \hat{x} < 1$ to $\psi(\overline{\theta}) = \exp(\rho\overline{\theta})$. Define similarly the function

$$\varphi(z) = \frac{1}{1-F(z)} \int_z^{\overline{\theta}} \theta f(\theta)\,\mathrm{d}\theta,$$

for any $z \in [\underline{\theta}, \overline{\theta}[$ and $\varphi(\overline{\theta}) = \overline{\theta}$. Again, it is not difficult to show that φ is a continuous and strictly increasing function from $\varphi(\underline{\theta}) = \hat{m} < 0$ to $\varphi(\overline{\theta}) = \overline{\theta}$.

Since ψ is strictly increasing, it thus suffices to show that $\psi(\theta^B) > 1 = \psi(\theta^W)$ to prove that $\theta^W < \theta^B$. Applying Jensen's inequality, it follows that

$$\psi(z) > \exp(\rho\varphi(z)),$$

for any $z \in]\underline{\theta}, \overline{\theta}[$, where the strict inequality comes from the strict convexity of $z \mapsto \exp(\rho z)$ and the non degeneracy of F. In particular, Jensen's inequality holds with equality at $\underline{\theta}$ and $\overline{\theta}$, but, by the intermediate value theorem, it must be that θ^B (as well as θ^W) lie in the open interval $]\underline{\theta}, 0[$. Thus, we have

$$\psi(\theta^B) > 1,$$

since $\varphi(\theta^B) = 0$ and $\theta^B \neq \underline{\theta}, \overline{\theta}$.

C. Mathematical appendix for chapter 3

C.1. Proofs for section 3.4.1

C.1.1. Proof of lemma 12

Let $\sigma \in \Sigma$ and suppose that there exist $\mu, \mu' \in \mathrm{supp}(\sigma)$ with $p(\mu) = p(\mu')$. Consider the following market:

$$\tilde{\mu} = \frac{\sigma(\mu)}{\sigma(\mu) + \sigma(\mu')} x + \frac{\sigma(\mu')}{\sigma(\mu) + \sigma(\mu')} x'.$$

By the convexity of $M_{p(\mu)}$, $p(\tilde{\mu}) = p(\mu)$. Define σ' in the following way: $\sigma'(\tilde{\mu}) = \sigma(\mu) + \sigma(\mu')$, $\sigma'(\mu) = \sigma'(\mu') = 0$ and $\sigma' = \sigma$ otherwise. Is it easy to check that $\sum_{\mu \in \mathrm{supp}(\sigma)} \sigma(\mu)\, W(\mu) = \sum_{\mu \in \mathrm{supp}(\sigma')} \sigma'(\mu)\, W(\mu)$. We can iterate this operation as many times as the number of pairs $\nu, \nu' \in \mathrm{supp}(\sigma')$ such that $p(\nu) = p(\nu')$ to finally obtain the desired conclusion.

C.1.2. Proof of lemma 13

Let μ^0 be an inefficient aggregate market, hence for any optimal segmentation $\sigma \in \Sigma(\mu^0)$, $|\mathrm{supp}(\sigma)| \geq 2$. Let σ be a direct and optimal segmentation of μ^0 and $\mu \in \mathrm{supp}(\sigma)$ such that μ is in the interior of $M_{p(\mu)}$. Let ν be any other market in the support of σ. Consider the market:

$$\xi = \frac{\sigma(\mu)}{\sigma(\mu) + \sigma(\nu)} \mu + \frac{\sigma(\nu)}{\sigma(\mu) + \sigma(\nu)} \nu.$$

Because μ^0 is inefficient, it is without loss of generality to assume that ξ is also inefficient.

Denote $\bar{\mu}$ (resp. $\bar{\nu}$) the projection of ξ on the boundary of the simplex M in direction of μ (resp. ν). For σ to be optimal, the segmentation of ξ between μ with probability $\sigma(\mu)/(\sigma(\mu)+\sigma(\nu))$ and ν with probability $\sigma(\nu)/(\sigma(\mu)+\sigma(\nu))$ must be optimal. In particular, it must be optimal among any segmentation on $[\bar{\mu}, \bar{\nu}]$.

There exists a one-to-one mapping $f\colon [\bar{\mu}, \bar{\nu}] \to [0, 1]$ such that for any

$\gamma \in [\bar{\mu}, \bar{\nu}]$, $\gamma = f(\gamma)\bar{\mu} + (1 - f(\gamma))\bar{\nu}$. Thus, the set $[\bar{\mu}, \bar{\nu}]$ can be seen as all the distributions on a binary set of states of the world $\{\bar{\mu}, \bar{\nu}\}$, where for any $\gamma \in [\bar{\mu}, \bar{\nu}]$, $f(\gamma)$ is the probability of $\bar{\mu}$.

Therefore, the maximization program,

$$\max_{\sigma \in \Sigma(\xi)} \sum_{\gamma \in \mathrm{supp}(\sigma)} \sigma(\gamma) W(\gamma) \tag{$\bar{\mathrm{S}}$}$$

where

$$\Sigma(\xi) := \left\{ \sigma \in \Delta\big([\bar{\mu}, \bar{\nu}]\big) \,\middle|\, \sum_{\gamma \in \mathrm{supp}(\sigma)} \sigma(\gamma)\gamma = \xi \right\},$$

is a Bayesian persuasion problem (Kamenica and Gentzkow, 2011), with a binary state of the world and a finite number of actions. Hence, applying theorem 1 in Lipnowski and Mathevet (2017), there exists an optimal segmentation only supported on extreme points of sets $M \in \mathcal{M}^{[\bar{\mu}, \bar{\nu}]} := \big\{ M_k \cap [\bar{\mu}, \bar{\nu}] \mid k \in \mathcal{K} \text{ and } M_k \cap [\bar{\mu}, \bar{\nu}] \neq \emptyset \big\}$. It happens that for any $M \in \mathcal{M}^{[\bar{\mu}, \bar{\nu}]}$, so that $M = M_k \cap [\bar{\mu}, \bar{\nu}]$ for some k, if γ is an extreme point of M, then it is on the boundary of (M_k).

Let (μ', ν') with respective probabilities $(\alpha, 1 - \alpha)$ be a solution to ($\bar{\mathrm{S}}$) where μ' and ν' are extreme points of some $M \in \mathcal{M}^{[\bar{\mu}, \bar{\nu}]}$. We now consider the segmentation $\bar{\sigma}$ such that $\bar{\sigma}(\gamma) = \sigma(\gamma)$ for all $\gamma \in \mathrm{supp}(\sigma) \setminus \{\mu, \nu\}$, $\bar{\sigma}(\mu') = (\sigma(\mu) + \sigma(\nu))\alpha$, $\bar{\sigma}(\nu') = (\sigma(\mu) + \sigma(\nu))(1 - \alpha)$, and $\bar{\sigma} = 0$ otherwise. One can easily check that $\bar{\sigma} \in \Sigma(\mu^0)$. If $\bar{\sigma}$ is not direct, that is, there exists $\gamma \in \mathrm{supp}(\bar{\sigma})$ such that (w.l.o.g.) $p(\gamma) = p(\mu')$, then construct a direct segmentation $\bar{\bar{\sigma}}$ following the same process as in the proof of lemma 12. Then, if $\bar{\bar{\sigma}}$ is not only supported on boundaries of sets $\{M_k\}_{k \in \mathcal{K}^0}$, reiterate the same process as above, until you reach the desired conclusion.

C.2. Proofs for section 3.4.2.

C.2.1. Proof for proposition 13

Fix an aggregate market μ^0 and let $\sigma \in \Sigma(\mu^0)$ be optimal and direct. Suppose by contradiction that there exist $\mu, \mu' \in \mathrm{supp}(\sigma)$ such that $v_a := \min\{\mathrm{supp}(\mu)\} < \max\{\mathrm{supp}(\mu')\} =: v_d$ and $v_b := \min\{\mathrm{supp}(\mu')\} < \max\{\mathrm{supp}(\mu)\} =: v_c$. Assume further, without loss of generality, that $\min\{\mathrm{supp}(\mu)\} < \min\{\mathrm{supp}(\mu')\}$.

Let us define

$$\bar{\mu} := \frac{\sigma(\mu)}{\sigma(\mu) + \sigma(\mu')}\mu + \frac{\sigma(\mu')}{\sigma(\mu) + \sigma(\mu')}\mu'.$$

If σ is optimal, then we have

$$V(\bar{\mu}) = \frac{\sigma(\mu)}{\sigma(\mu)+\sigma(\mu')}W(\mu) + \frac{\sigma(\mu')}{\sigma(\mu)+\sigma(\mu')}W(\mu').$$

The proof consists in showing that we can improve on this splitting of $\bar{\mu}$ and thus obtains a contradiction.

Let $\varepsilon > 0$, and define $\check{\mu}, \check{\mu}'$ as follows:

$$\check{\mu}_k = \begin{cases} \mu_k + \varepsilon & \text{if} k = b \\ \mu_k - \varepsilon & \text{if} k = c \\ \mu_k & \text{otherwise,} \end{cases}$$

and,

$$\check{\mu}'_k = \begin{cases} \mu'_k - \dfrac{\sigma(\mu)}{\sigma(\mu')}\varepsilon & \text{if} k = b \\[2ex] \mu'_k + \dfrac{\sigma(\mu)}{\sigma(\mu')}\varepsilon & \text{if} k = c \\[2ex] \mu'_k & \text{otherwise.} \end{cases}$$

By construction, we have

$$\bar{\mu} = \frac{\sigma(\mu)}{\sigma(\mu)+\sigma(\mu')}\check{\mu} + \frac{\sigma(\mu')}{\sigma(\mu)+\sigma(\mu')}\check{\mu}'.$$

Note that v_a is still an optimal price for $\check{\mu}$. Indeed, for any $v_a \leq v_k \leq v_b$, the profit made by fixing price v_k is equal in markets μ and $\check{\mu}$ and for any $v_b < v_k \leq v_c$ the profit made by fixing price v_k is strictly lower in $\check{\mu}$ than in μ. On the contrary, $\phi(\check{\mu}') \geq \phi(\mu')$ and it is possible that the inequality holds strictly. In any case, it must be that $\phi(\check{\mu}') = v_e$ for $b \leq e \leq d$. Denote $\alpha := \sigma(\mu)/(\sigma(\mu)+\sigma(\mu'))$, hence $\sigma(\mu)/\sigma(\mu') = \alpha/(1-\alpha)$. Remark that

$$\begin{aligned} &\alpha W(\check{\mu}) + (1-\alpha)W(\check{\mu}') - \big(\alpha W(\mu) + (1-\alpha)W(\mu')\big) \\ &\qquad\qquad = \alpha\big(W(\check{\mu}) - W(\mu)\big) + (1-\alpha)\big(W(\check{\mu}') - W(\mu')\big), \end{aligned}$$

that

$$W(\check{\mu}) - W(\mu) = \varepsilon\big(\lambda_b(v_b - v_a) - \lambda_c(v_c - v_a)\big),$$

and that

$$W(\breve{\mu}')-W(\mu') = \left(-\sum_{k>e}\lambda_k\mu'_k(v_e-v_b) - \sum_{b<k\leq e}\lambda_k\mu'_k(v_k-v_b) + \lambda_c\left(\frac{\alpha}{1-\alpha}\right)\varepsilon(v_c-v_e)\right).$$

Hence, rearranging terms, we obtain:

$$\begin{aligned}
&\alpha W(\breve{\mu}) + (1-\alpha)W(\breve{\mu}') - \big(\alpha W(\mu) + (1-\alpha)W(\mu')\big) = \\
&\qquad \alpha\varepsilon\big(\lambda_b(v_b-v_a) - \lambda_c(v_c-v_a)\big) \\
&+(1-\alpha)\left(-\sum_{k>e}\lambda_k\mu'_k(v_e-v_b) - \sum_{b<k\leq e}\lambda_k\mu'_k(v_k-v_b) + \lambda_c\left(\frac{\alpha}{1-\alpha}\right)\varepsilon(v_c-v_e)\right)
\end{aligned}$$

which simplifies into

$$\begin{aligned}
&\alpha W(\breve{\mu}) + (1-\alpha)W(\breve{\mu}') - \big(\alpha W(\mu) + (1-\alpha)W(\mu')\big) = \\
&\alpha\varepsilon\lambda_b(v_b-v_a) - \alpha\varepsilon\lambda_c(v_e-v_a) - (1-\alpha)\left(\sum_{k>e}\lambda_k\mu'_k(v_e-v_b) + \sum_{b<k\leq e}\lambda_k\mu'_k(v_k-v_b)\right)
\end{aligned}$$

But remark that

$$\begin{aligned}
&\alpha\varepsilon\lambda_b(v_b-v_a) - \alpha\varepsilon\lambda_c(v_e-v_a) - (1-\alpha)\left(\sum_{k>e}\lambda_k\mu'_k(v_e-v_b) + \sum_{b<k\leq e}\lambda_k\mu'_k(v_k-v_b)\right) \\
&>\alpha\varepsilon\lambda_b(v_b-v_a) - \alpha\varepsilon\lambda_{b+1}(v_e-v_a) - (1-\alpha)\left(\sum_{k>e}\lambda_{b+1}\mu'_k(v_e-v_b) + \sum_{b<k\leq e}\lambda_{b+1}\mu'_k(v_k-v_b)\right) \\
&=\alpha\varepsilon\lambda_b(v_b-v_a) - \lambda_{b+1}\left[\alpha\varepsilon(v_e-v_a) - (1-\alpha)\left(\sum_{k>e}\mu'_k(v_e-v_b) + \sum_{b<k\leq e}\mu'_k(v_k-v_b)\right)\right]
\end{aligned} \tag{C.1}$$

Finally,

$$\text{(C.1)} \geq 0 \iff \frac{\lambda_b}{\lambda_{b+1}} \geq \kappa$$

where

$$\kappa = \frac{\alpha\varepsilon(v_e-v_a) - (1-\alpha)\big(\sum_{k>e}\mu'_k(v_e-v_b) + \sum_{b<k\leq e}\mu'_k(v_k-v_b)\big)}{\alpha\varepsilon(v_b-v_a)}$$

which ends the proof.

C.3. Proofs of Section 3.4.3.

C.3.1. Proof of proposition 14

As argued in the core of the text, all markets with uniform price v_u belonging to no-rent region must be optimally segmented by splitting μ^* between

$$\mu^s = \left(\frac{\mu_1^*}{\sigma}, \frac{\mu_2^*}{\sigma}, \ldots, \mu_u^s, 0, \ldots, 0\right)$$

and

$$\mu^r = \left(0, 0, \ldots, \mu_u^r, \frac{\mu_{u+1}^*}{1-\sigma}, \ldots, \frac{\mu_K^*}{1-\sigma}\right).$$

Such a segmentation indeed gives no rents to the monopolist if v_u is an optimal price in both μ^s and μ^r. That is, if:

$$v_1 = v_u \mu_u^s \geq v_j \left(\sum_{i=j}^{u-1} \frac{\mu_i^*}{\sigma} + \mu_u^s\right), \quad \forall\, 2 \leq j \leq u-1 \tag{NR-s}$$

and

$$v_u \geq v_j \left(\sum_{i=j}^{K} \frac{\mu_i^*}{1-\sigma}\right), \quad \forall\, u+1 \leq j \leq K \tag{NR-r}$$

As such, any optimal segmentation under strong redistributive preferences that maximizes consumer surplus must have

$$\mu_u^s = \frac{v_1}{v_u},$$

and

$$\sigma = \frac{v_u}{v_u - v_1} \sum_{i=1}^{u-1} \mu_i^*,$$

as well as

$$\mu_u^r = \frac{\mu_u^* v_u - \sum_{i=1}^{u} \mu_i^* v_1}{\sum_{i=u}^{K} \mu_i^* v_u - v_1}.$$

These conditions pin down the segmentation σ^{NR}. Conditions (NR-s) and (NR-r) are satisfied whenever σ^{NR} is efficient, which concludes the proof.

It is also interesting to note that conditions (NR-s) and (NR-r) define the no-rent region inside M_u as a convex polytope. Indeed, we can rearrange both conditions

and get:

$$0 \geq -\alpha_j \sum_{i=1}^{j-1} \mu_i^* + \left(1 - \alpha_j\right) \sum_{i=j}^{u-1} \mu_i^* \qquad \forall\, 2 \leq j \leq u-1 \tag{NR-s}$$

and

$$-\frac{v_1}{v_j(v_u - v_1)} \geq -\beta_j \sum_{i=u}^{j-1} \mu_i^* + \left(1 - \beta_j\right) \sum_{i=j}^{K} \mu_i^* \qquad \forall\, u+1 \leq j \leq K \tag{NR-r}$$

where

$$\alpha_j = \frac{v_1(v_u - v_j)}{v_j(v_u - v_1)},$$

and

$$\beta_j = \frac{v_u^2}{v_j(v_u - v_1)}.$$

The conditions expressed above define $K-2$ half-spaces in $\mathbb{R}^K$. The no-rent region in M_u is thus given by the closed polytope defined by the intersection of such half-spaces. We can represent such polytope as follows:

$$\Gamma_u = \left\{\mu \in M_u \,|\, \mathbf{A} \cdot \mu \leq \mathbf{z}\right\},$$

with

$$\mathbf{A} = \begin{pmatrix} \mathbf{S} & \mathbf{O_S} \\ \mathbf{O_R} & \mathbf{R} \end{pmatrix} \in \mathbb{R}^{(K-2)\times K}$$

and

$$\mathbf{z} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -\frac{v_1}{v_{u+1}(v_u - v_1)} \\ \vdots \\ -\frac{v_1}{v_K(v_u - v_1)} \end{pmatrix} \in \mathbb{R}^{K-2}$$

where $\mathbf{O_S}$ and $\mathbf{O_R}$ are null matrices with respective dimensions $(u-2) \times (u-1)$

and $(K-u)\times(K+1-u)$, and where

$$\mathbf{S}=\begin{pmatrix}-\alpha_2 & 1-\alpha_2 & \cdots & 1-\alpha_2 & 1-\alpha_2\\ -\alpha_3 & -\alpha_3 & \cdots & 1-\alpha_3 & 1-\alpha_3\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ -\alpha_{u-2} & -\alpha_{u-2} & \cdots & 1-\alpha_{u-2} & 1-\alpha_{u-2}\\ -\alpha_{u-1} & -\alpha_{u-1} & \cdots & -\alpha_{u-1} & 1-\alpha_{u-1}\end{pmatrix}\in\mathbb{R}^{(u-2)\times(u-1)},$$

and

$$\mathbf{R}=\begin{pmatrix}-\beta_{u+1} & 1-\beta_{u+1} & \cdots & 1-\beta_{u+1} & 1-\beta_{u+1}\\ -\beta_{u+2} & -\beta_{u+2} & \cdots & 1-\beta_{u+2} & 1-\beta_{u+2}\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ -\beta_{K-1} & -\beta_{K-1} & \cdots & 1-\beta_{K-1} & 1-\beta_{K-1}\\ -\beta_K & -\beta_K & \cdots & -\beta_K & 1-\beta_K\end{pmatrix}\in\mathbb{R}^{(K-u)\times(K+1-u)}.$$

Bibliography

Aliprantis, C. D. and K. C. Border (2006): *Infinite Dimensional Analysis*, Berlin Heidelberg: Springer-Verlag, third ed.

Clarke, F. (2013): *Functional Analysis, Calculus of Variations and Optimal Control*, vol. 264 of *Graduate Texts in Mathematics*, London: Springer-Verlag.

Dupuis, P. and R. S. Ellis (1997): *A Weak Convergence Approach to the Theory of Large Deviations*, Wiley.

Kamenica, E. and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101, 2590–2615.

Lipnowski, E. and L. Mathevet (2017): "Simplifying Bayesian Persuasion," *Working Paper*.

Panik, M. J. (1993): *Fundamentals of Convex Analysis*, Dordrecht: Springer Netherlands.

Rochet, J.-C. (1987): "A necessary and sufficient condition for rationalizability in a quasi-linear context," *Journal of Mathematical Economics*, 16, 191–200.

Rockafellar, R. T. (1970): *Convex Analysis*, Princeton, New Jersey: Princeton University Press.

Royden, H. and P. Fitzpatrick (2010): *Real Analysis*, Pearson, fourth ed.